Cloud Computing Concepts and Experiences

Sebastien Goasguen –sebgoa@clemson.edu

School of Computing

Clemson University, Clemson, SC

Scientific Associate at CERN

Summer 2009 and Summer 2010

Outline

- **Cloud Basics**
- Building a Cloud Provider
  - Lxcloud @ CERN
- VOCs and Clouds
  - Research done at Clemson

What is Cloud Computing ?

A few references

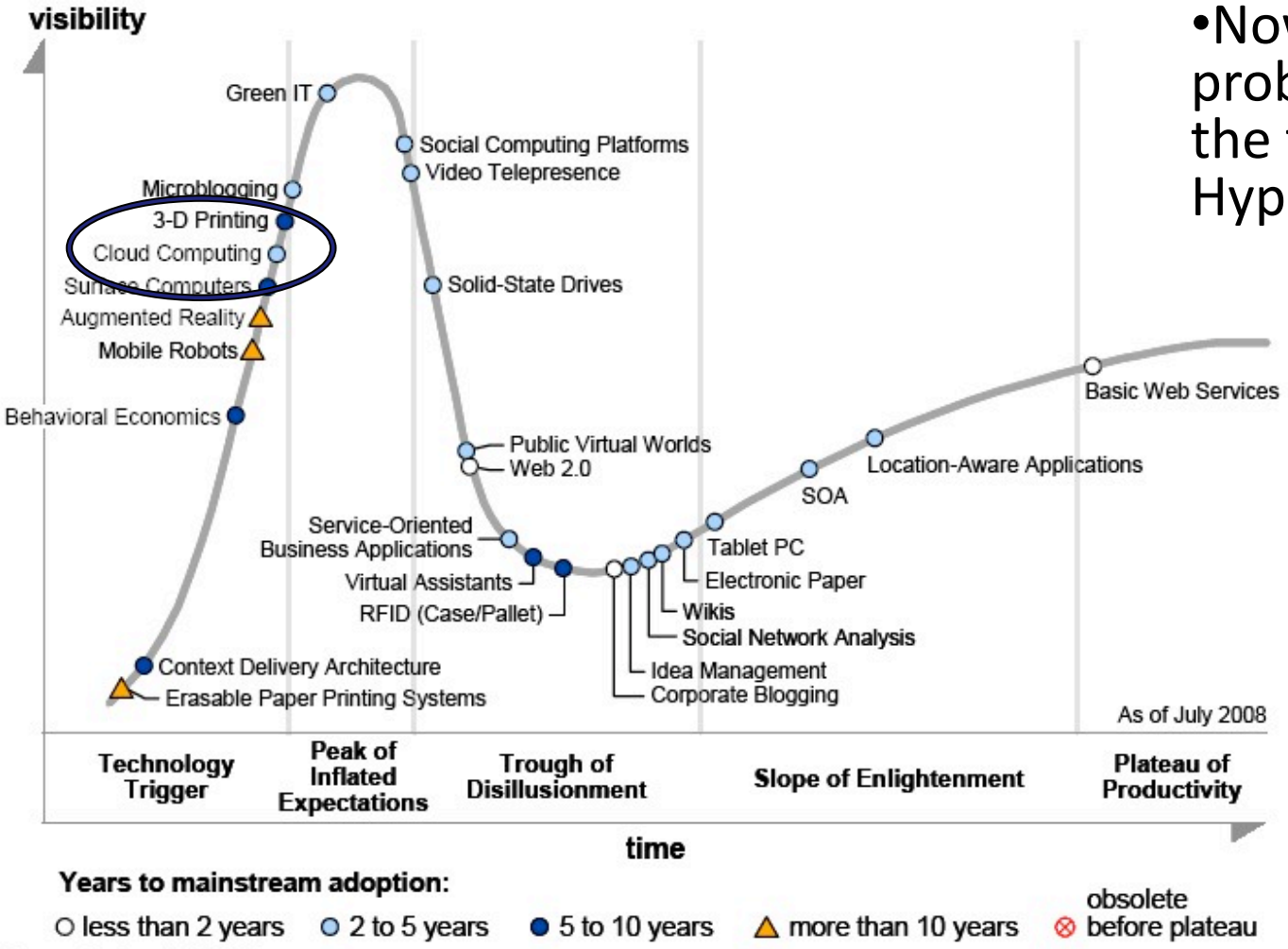
"Above the clouds: A Berkeley view of cloud computing"
http://berkeleyclouds.blogspot.com/

"A break in the clouds: towards a cloud definition"
L.M Vaquero et al. SIGCOMM computer communication review, 2008.
http://portal.acm.org/citation.cfm?id=1496100

"An *EGEE* Comparative Study - Grid *cloud* comparative study"
M-Elian Begin, 2009

On the Hype curve

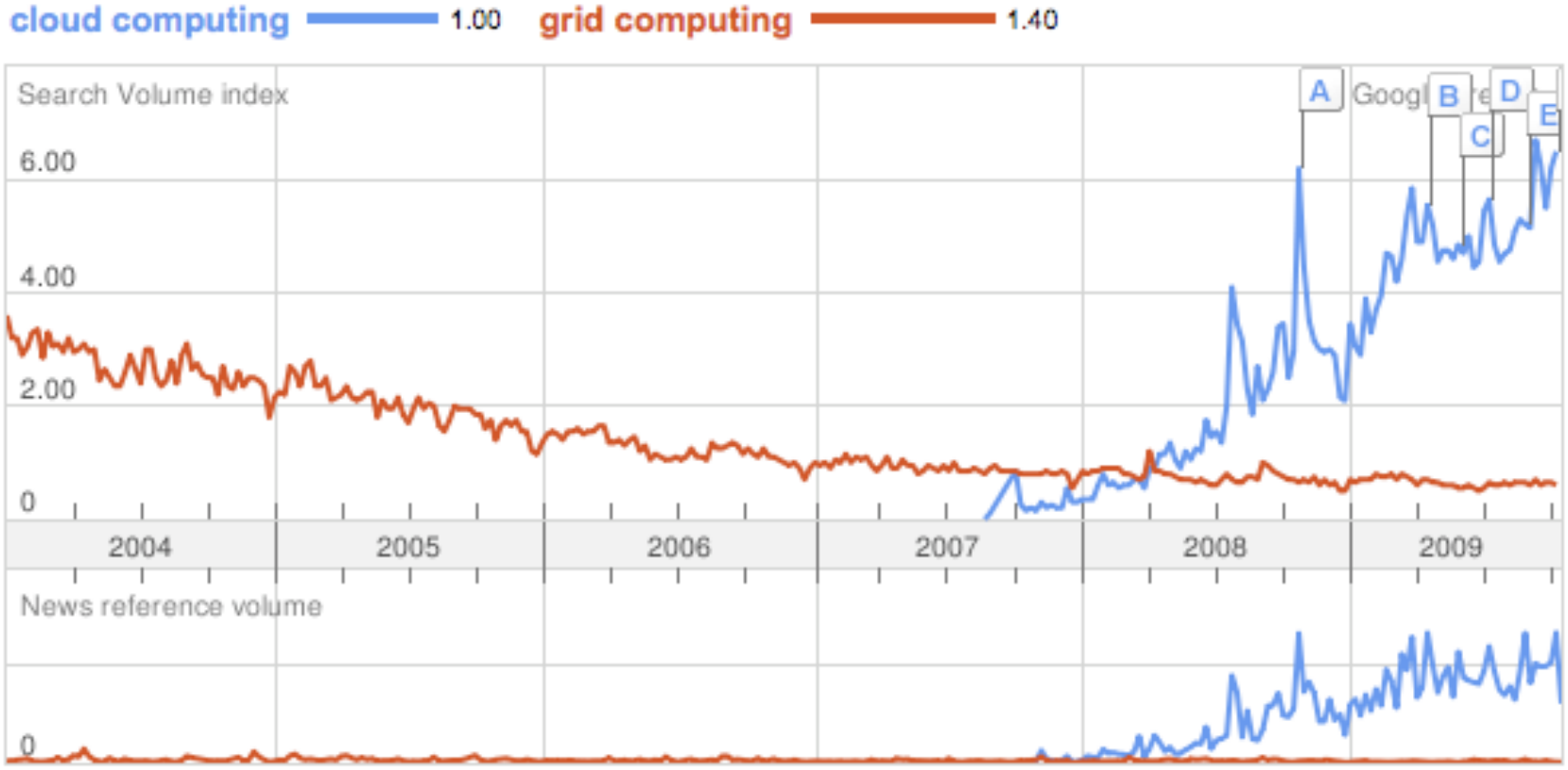
Figure 1. Hype Cycle for Emerging Technologies, 2008



Source: Gartner (July 2008)

- Now probably at the top of the Hype –Oct 09

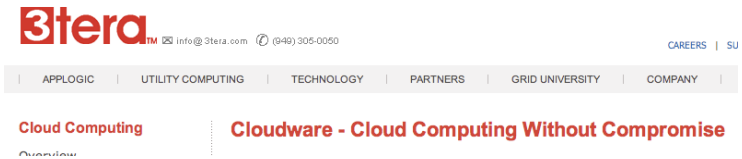
Trendy…



•Source: http://www.google.com/trends

Cloud formation

Amazon Elastic Compute Cloud (Amazon EC2) - *Beta*



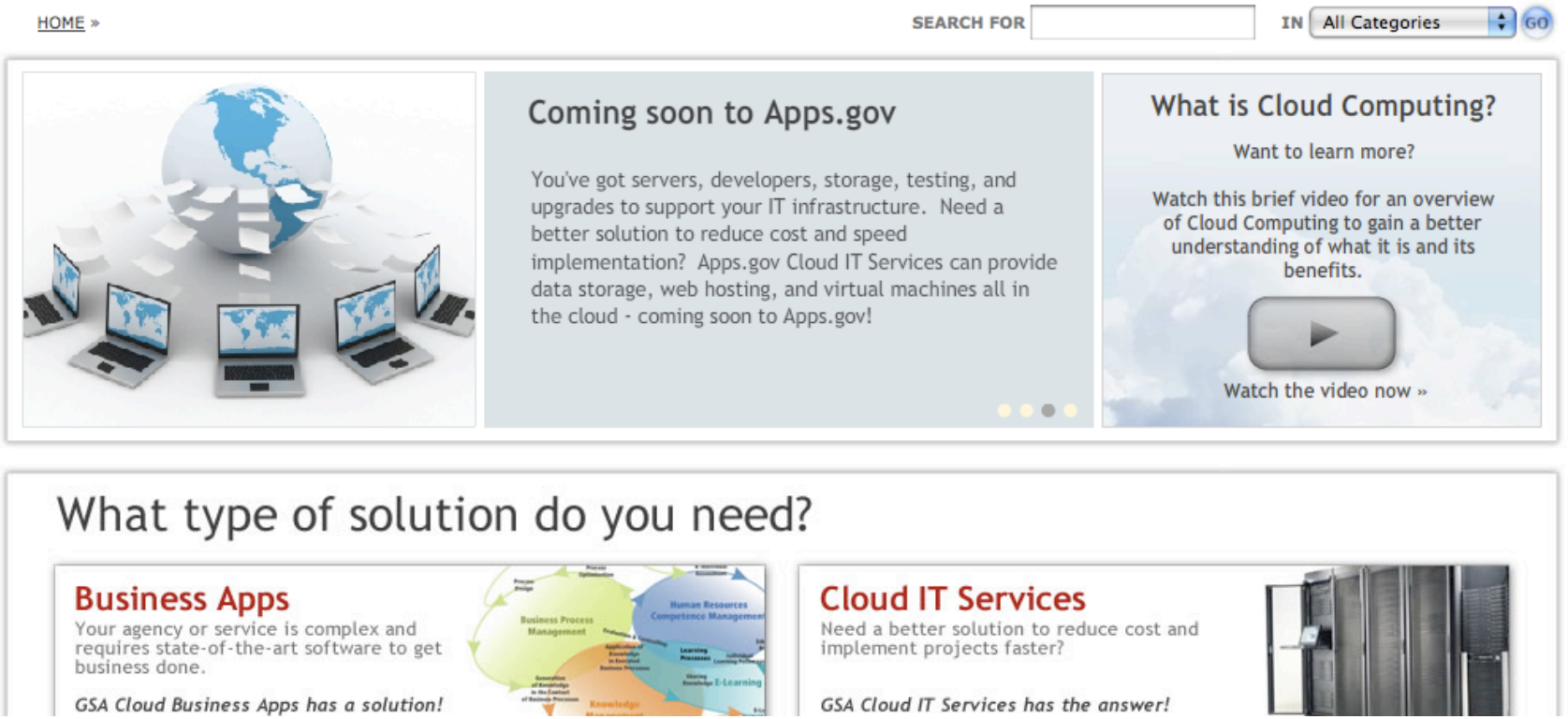
MOSSO
the hosting cloud

the globus project
www.globus.org

•**Slide adapted from Rich Wolski, UCSB**

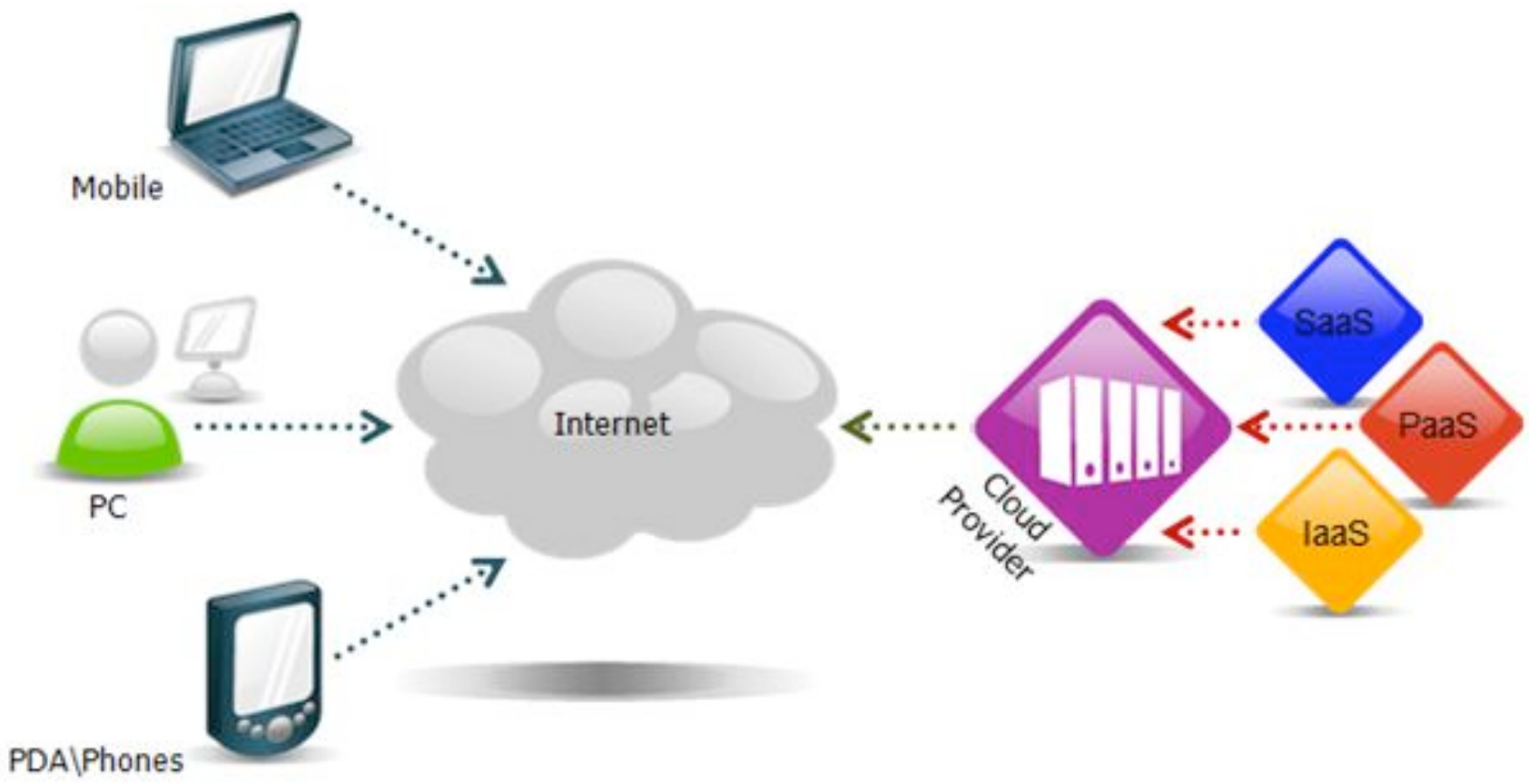
White House is going to the Cloud

- Reduce costs … See Apps.gov



DOE and NASA too (Check novacc.org)

Everything is moving to the Cloud…
Stay on Earth though !



•http://contactdubai.com/tag/saas-software-or-storage-as-a-service

An "Old" idea: OSI/Anatomy of the Grid/ Windows architectures…

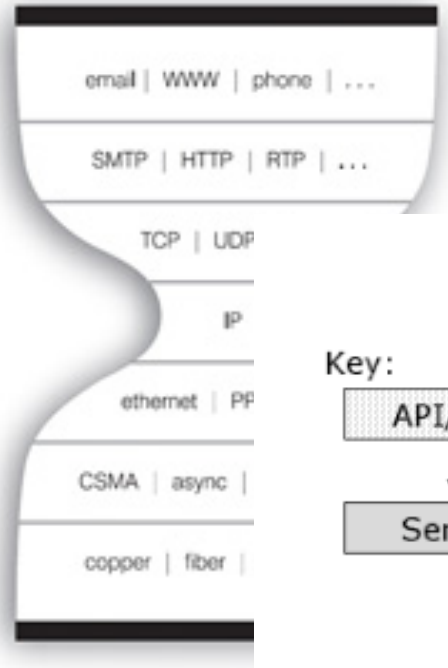
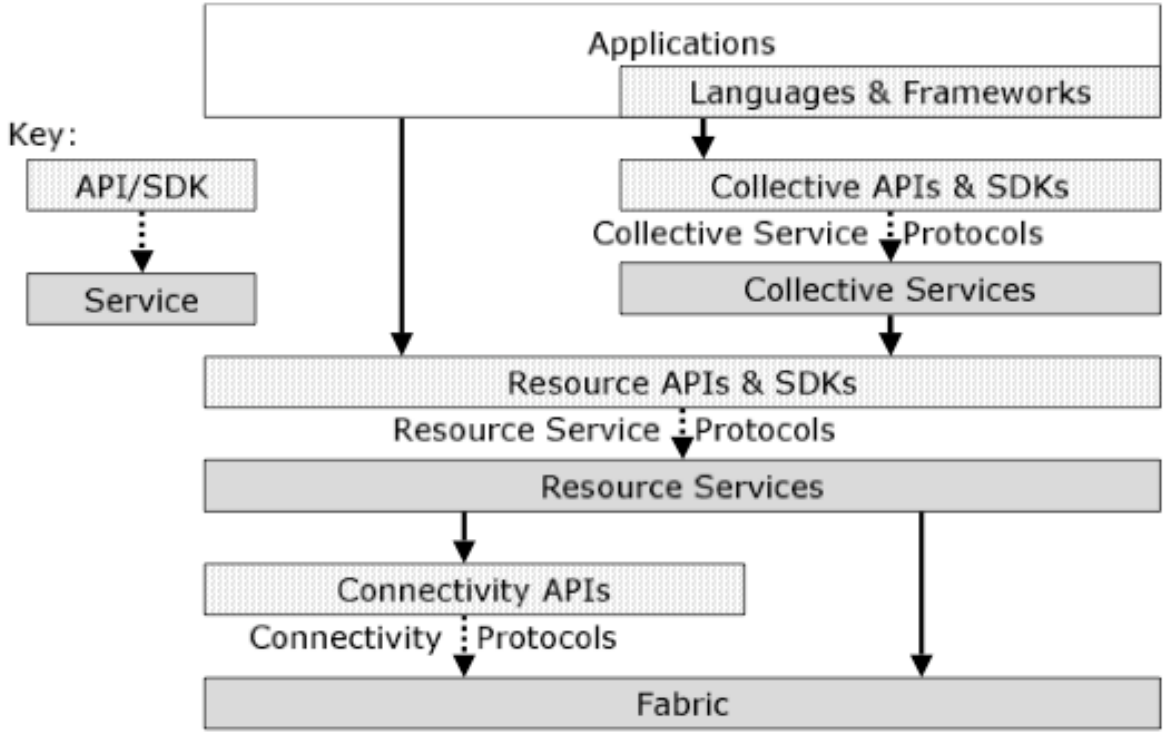


Figure 4.1 Hourglass architecture of the Internet

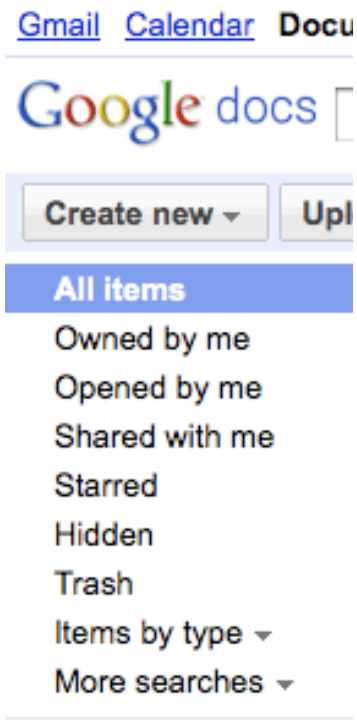


What is the Cloud ? The *aaS

- **SaaS –Software as a Service -**
- **PaaS –Platform as a Service -**
- **Iaas –Infrastructure as a Service –**
- Service composition at all layers of distributed system. Builds a system of system
- Software and hardware reuse

- Tendency for the *aaS-itus but these three are the main ones

Software as a Service



- Google Analytics Data Export API
- Google Apps APIs
  - Getting Started
  - List of All Apps APIs
  - Developer's Guides:
    - Admin Settings API
    - Provisioning API
    - Email Migration API
    - Domain Shared Contacts API
    - User Profiles API
    - Email Settings API
  - 2-Legged OAuth for Data API Feeds
- Google Base Data API
- Blogger Data API
- Google Booksearch Data API
- Google Calendar Data API
- Google Finance Portfolio Data API
- Google Health Data API
- Google Maps Data API
- Picasa Web Albums Data API
- Google Sidewiki Data API
- Google Sites Data API
- Google Spreadsheets Data API
- Google Webmaster Tools Data API
- YouTube Data API

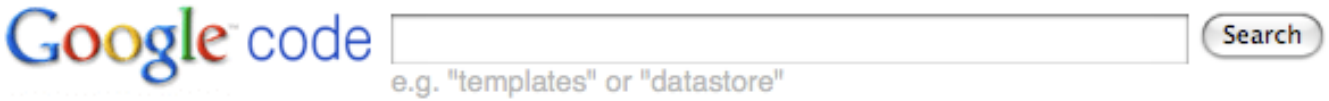
Sky is the limit…

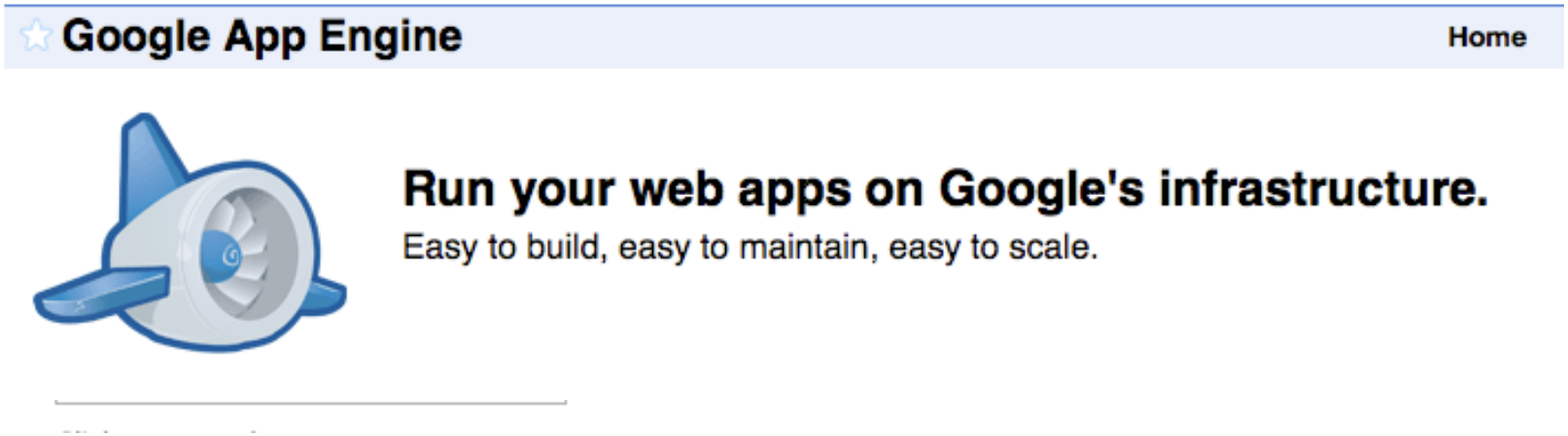
- Phone apps…FermiVoice ?

Platform as a Service

Windows® Azure is a cloud services operating system that serves as the **development, service hosting and service management environment** for the Windows Azure platform. Windows Azure provides developers with on-demand compute and storage to host, scale, and manage web applications on the internet through Microsoft® datacenters.

Windows Azure is an open platform that supports Microsoft and non-Microsoft languages and





Infrastructure as a Service / Coming of age of virtualization

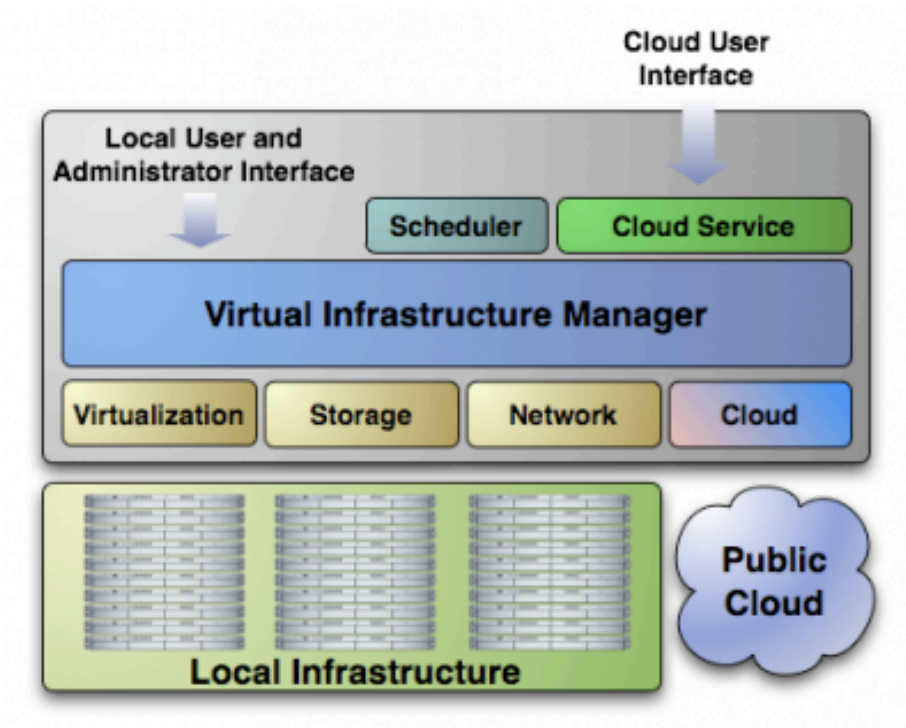
Sun to
farm
Networking

Infrastruct

- Amazon El
  Cloud (Am
- Amazon Sir
- Amazon Sir
  Service (An
- Amazon Clo
- Amazon Sir
  Service (An
- Amazon Ela
- AWS Premi

The Open Source Toolkit to Build your Cloud

OpenNebula is an **open and flexible tool that fits into existing data center environments** to **build any type of Cloud deployment**. OpenNebula can be primarily used as a virtualization tool to manage your virtual infrastructure in the data-center or cluster, which is usually referred as **Private Cloud**. OpenNebula supports **Hybrid Cloud** to combine local infrastructure with public cloud-based infrastructure, enabling highly scalable hosting environments. OpenNebula also supports **Public Clouds** by providing Cloud interfaces to expose its functionality for virtual machine, storage and network management.



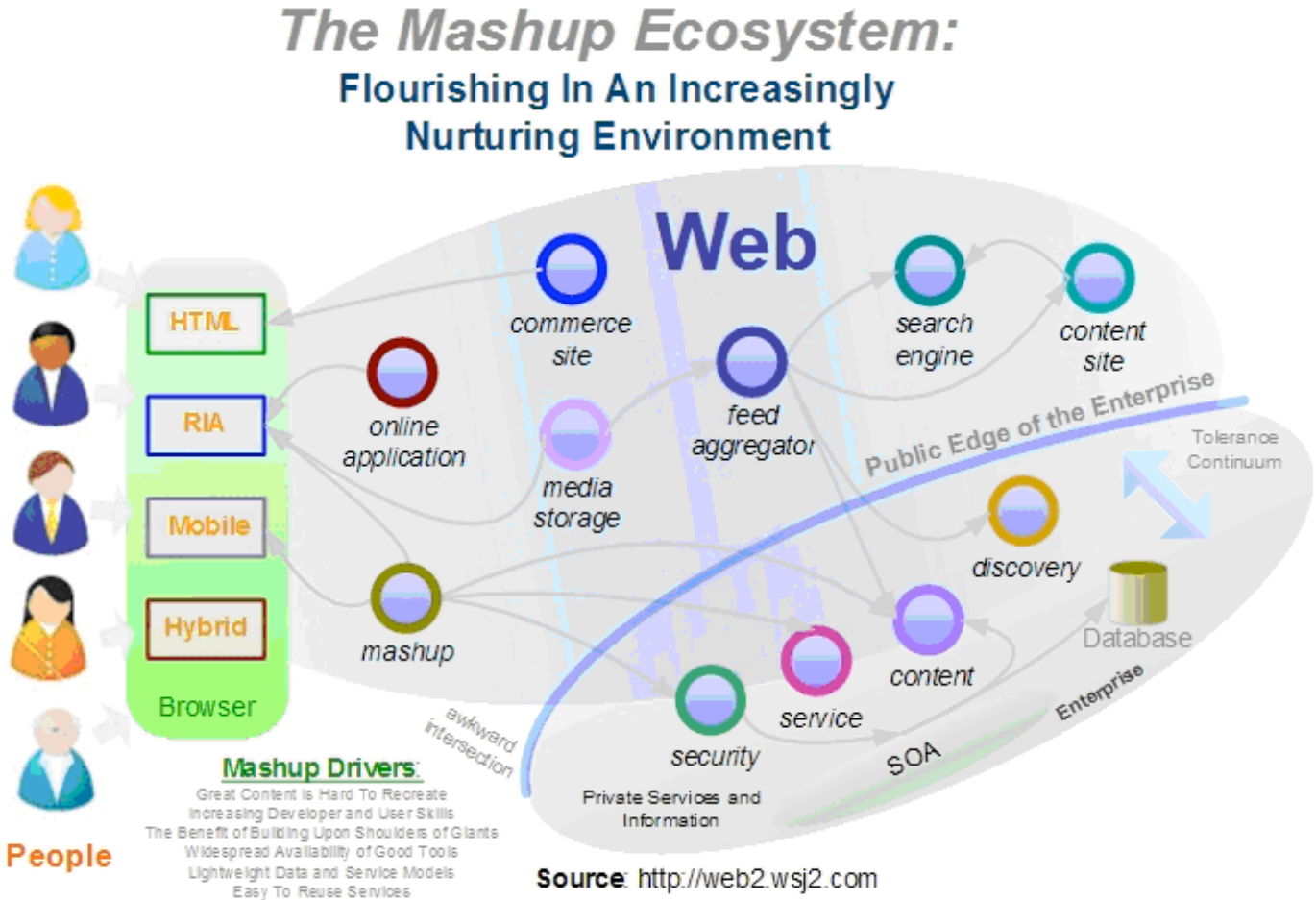
What is the Cloud ? The *aaS

- **SaaS –Software as a Service –**
  - Easy Access to hosted applications over the network. Most likely using your Browser
  - API to these applications
- **PaaS –Platform as a Service –**
  - Environment to deploy new applications
  - Restricted capabilities offered
  - API to this platform and access to SaaS API
- **Iaas –Infrastructure as a Service –**
  - Access to Hardware resources
  - API to make resource allocation requests

Key Features

- **You don't know what's behind but it works**
  - Transparency
- **You Pay what you use**
  - Utility pricing
- **You get what you ask for (On-demand)**
  - Read the fine prints of the SLAs…
- **It scales if you need more**
  - How far does it scale ?
  - Doesn't this mean the underlying resources are underutilized ?

Why now ?
Evolution of the Mashup Revolution thanks to an API "explosion"

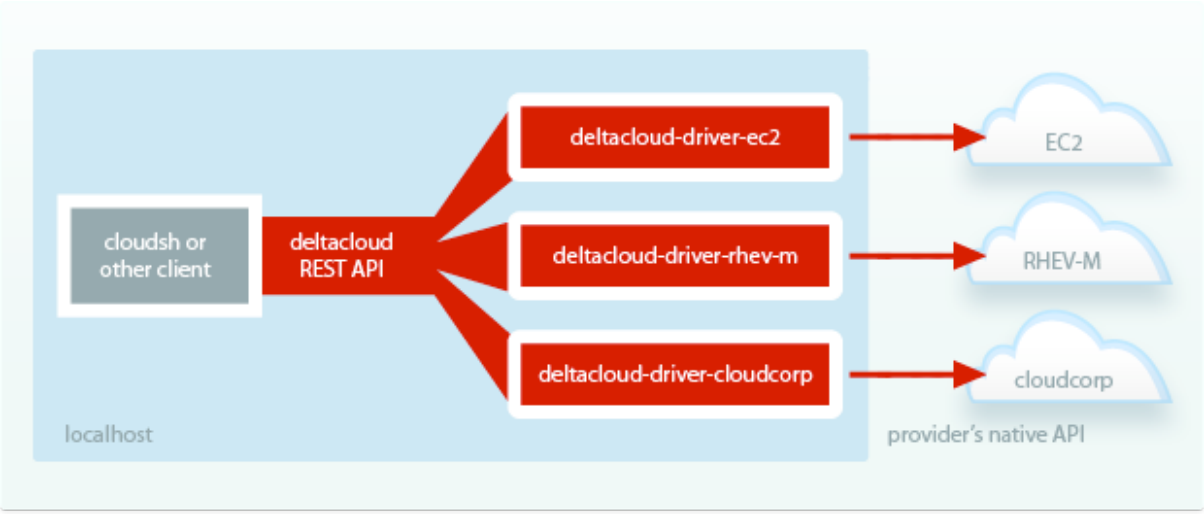


Why now ?

- Big Internet companies faced a **lot of data to analyze**: web logs…
- **Developed in house:** New file system (Hadoop), new analysis framework (Map-Reduce)
- **Massive amount of resources** all across the planet: >500,000 cores for Google ?
- Higher needs to consolidate: virtualization, energy costs.
- **New devices:** iPhone / G1
- A truly inter-connected planet

A few interesting things…to stir the pot

- Industry is leading. Is academia behind ?
- Who cares about standards ? (>20 bodies working on cloud standards…)
- We should switch paradigm and rewrite applications once they are 6 months old.



Outline

- Cloud Basics
- **Building a Cloud Provider**
  - **Lxcloud @ CERN (In collaboration with Ulrich Schwickerath, Ewan Roche, Belmiro Moreira and Romain Wartel)**
- VOCs and Clouds
  - Research done at Clemson

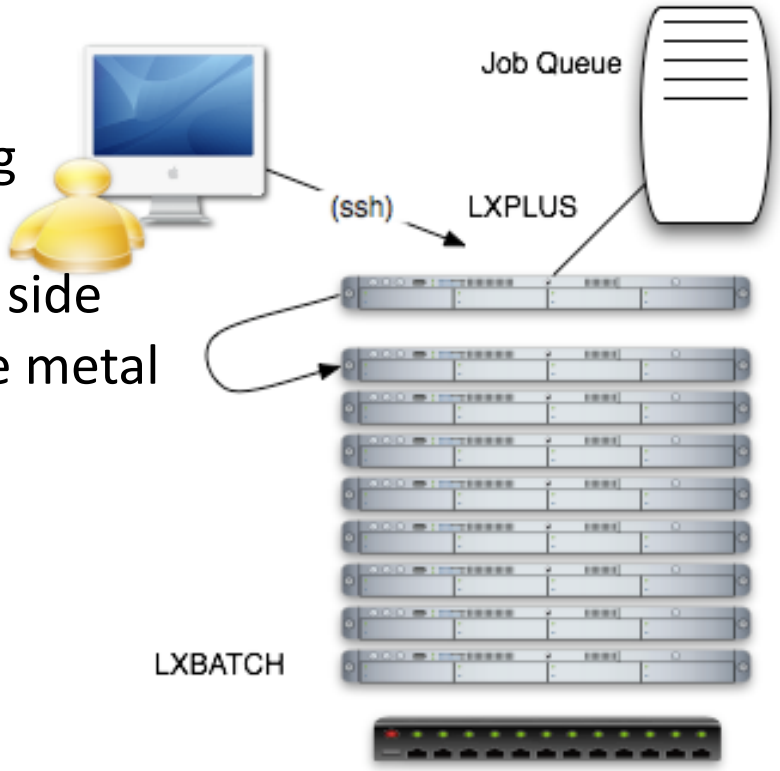
IaaS level

- For consolidating services
  - Used in IT for a while now
  - Fermi Grid services running in Xen VMs
- **For offering on-demand services**
  - **E.g VOBoxes, replace hardware request**
- **For virtualizing large scale services**
  - **Clusters on-demand**

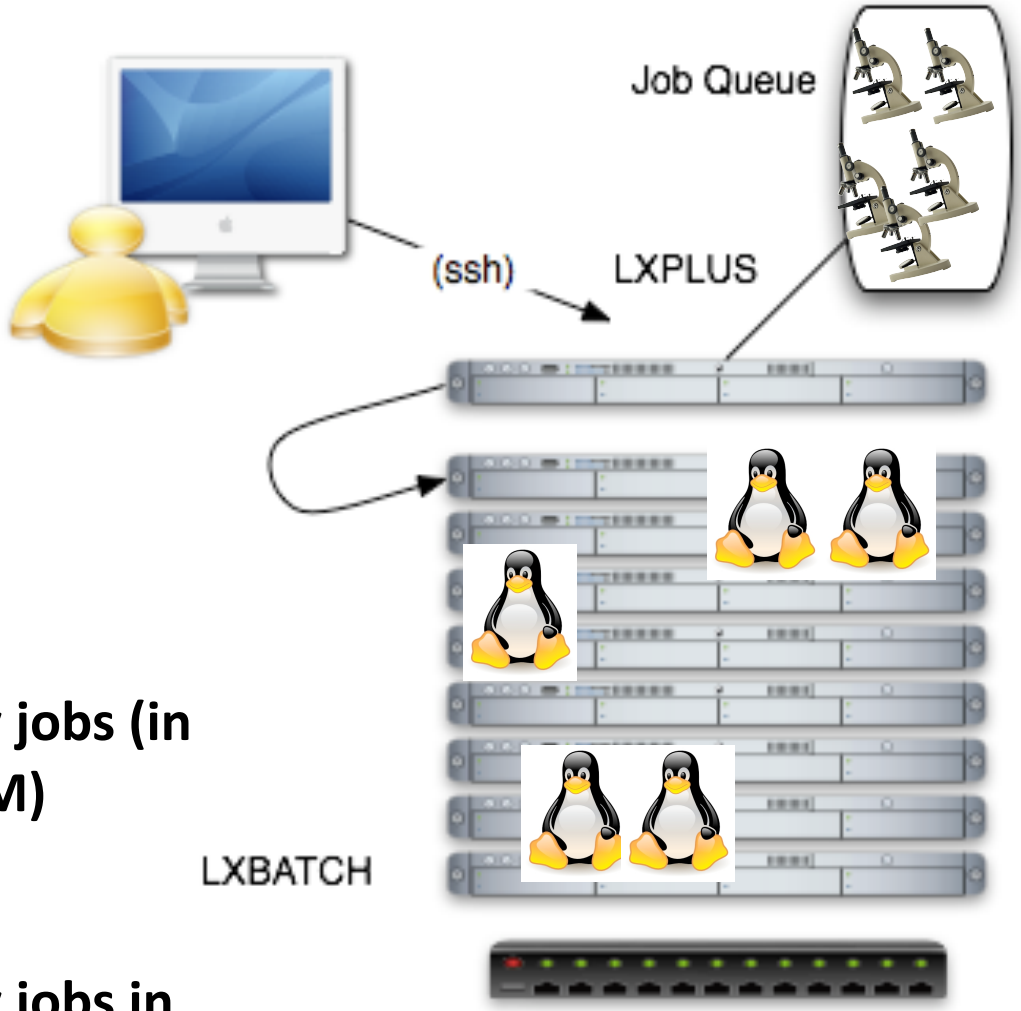
- Virtualization is a key enabler for IaaS

Why virtualizing "Batch" ?

- Run batch jobs within Virtual Machines
  - Better application environment
    - Custom made by user
  - Increased security
  - Better control on resource sharing
    - Multi-core apps
  - Increased flexibility on the admin side
    - Can run a preferred OS on the metal



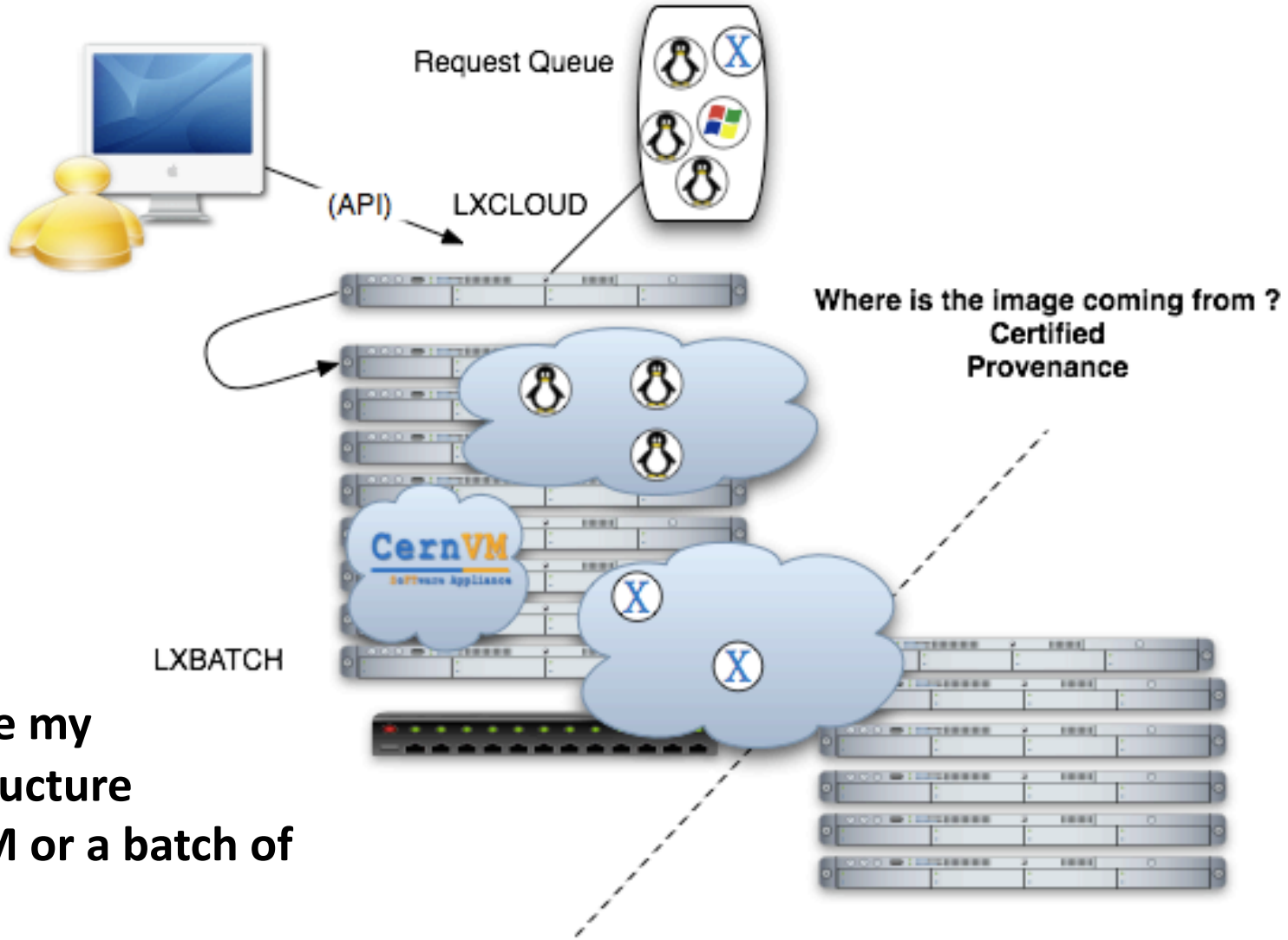
How to virtualize “Batch”…smoothly ?



Type 1:
Run my jobs (in your VM)

Type 2:
Run my jobs in my VM

Moving to the cloud:



Type 3:
Give me my infrastructure
i.e a VM or a batch of VMs

Deployment Models

Innovation in Cloud Computing Architectures

Model	Definition	Examples of Deployment
Private	Infrastructure is owned by a single organization and made available only to the organization	• Optimize and simplify **internal operation** • **SaaS/PaaS** support • IT consolidation within **large organizations** (Goverment Clouds, University Clouds…)
Public	Infrastructure is owned by a single organization and made available to other organizations	• **Commercial cloud providers** • **Community public clouds** by ICT service centers to enable scientific and educational projects to experiment with cloud computing • **Special purpose clouds** with dedicated capabilities (Science Clouds, HPC Clouds..) • **Regional clouds** to address regulatory or latency issues
Hybrid	Infrastructure is a composition of two or more clouds	• **Cloudbursting** to address peak demands • **Cloud Federation** to share infrastructure with partners • **Cloud Aggregation** to provide a larger resource infrastructure

dsa-research.org

Main components/characteristics

Set of Hypervisors

- Physical machines with a virtual machine monitor
- Xen or KVM ...or Hyper-V...or VMware ESx...

VM provisioning system

- OpenNebula
- Nimbus
- Eucalyptus
- Platform ISF
- or even traditional schedulers like PBS/Maui.

Image distribution mechanism

- Shared file system (e.g NFS, AFS, PVFS, Lustre...)
- Copy images (e.g scp, wget, Bittorent)

Networking

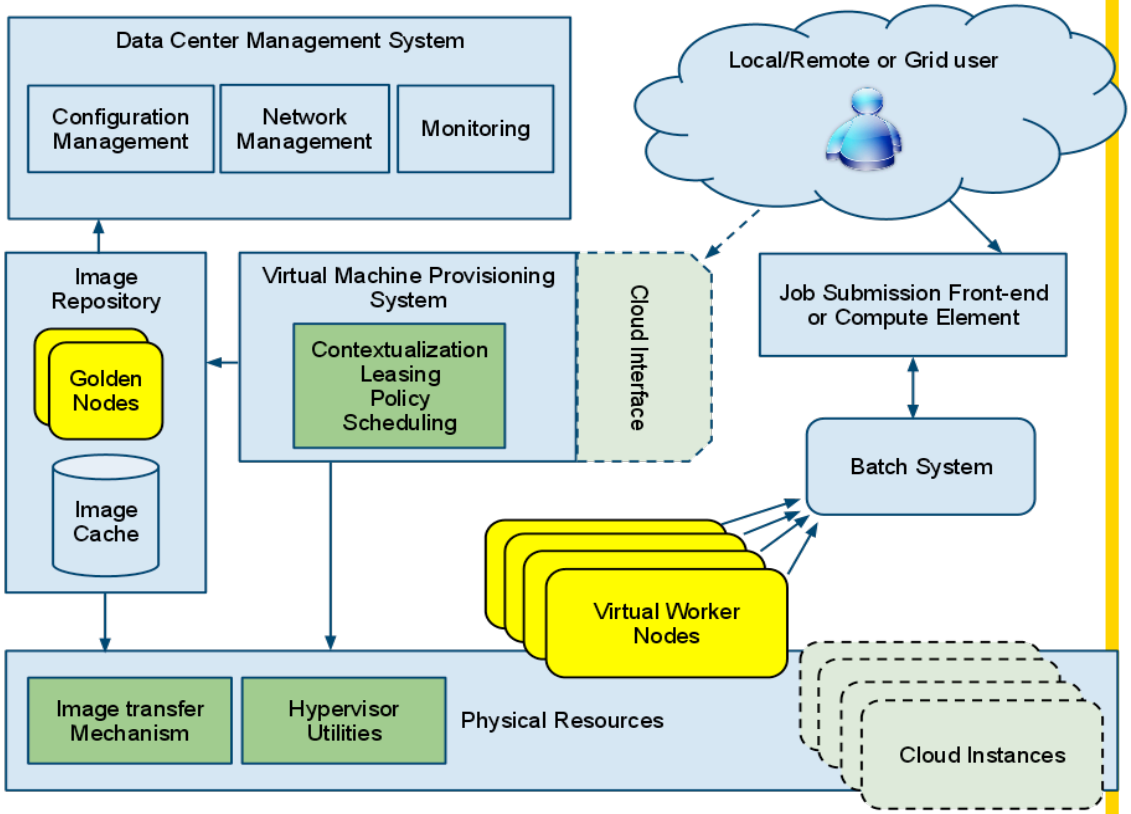
- Private / Public bridged
- NAT

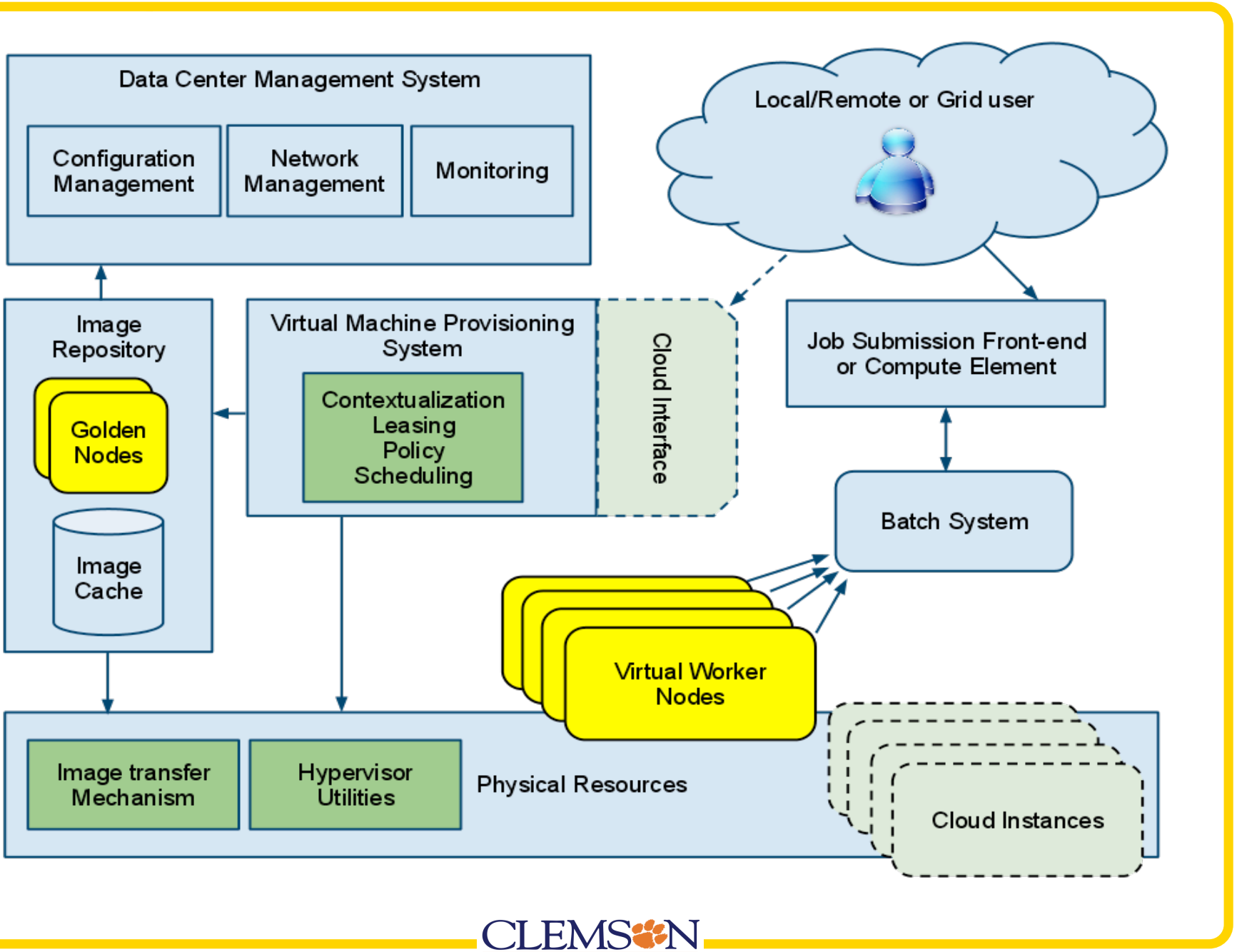
Thoughts for OSG…to stir the pot again

- Sites need to have hypervisors, that's a starting point. Without it/them there won't be OSG clouds.
- What VMM/Hypervisor they use does not matter…but my guess is that 80% will use KVM
- What provisioning system they use is a matter of local technical setup, taste and relationships
- Sites can do this now
- The hard problem is in the image transfer and trust…See HEPiX virtualization working group

CERN's LXCLOUD architecture

- Image repository with Golden nodes.
- VM instances not quattor managed have finite lifetime
- Specific IP/MACs are pinned to hypervisors
- Currently testing two provisioning system: Opennebula and Platform ISF.




Data Center Management System
Configuration Management
Network Management
Monitoring
Local/Remote or Grid user
Image Repository
Golden Nodes
Image Cache
Virtual Machine Provisioning System
Contextualization
Leasing
Policy
Scheduling
Cloud Interface
Job Submission Front-end or Compute Element
Batch System
Virtual Worker Nodes
Image transfer Mechanism
Hypervisor Utilities
Physical Resources
Cloud Instances

Provisioning system

OpenNebula and Platform ISF are currently being evaluated. Results shown in this talk were obtained with OpenNebula.

OpenNebula out of the University Compultense of Madrid

- C/C++ core with Ruby drivers and command line interface
- Mysql and Sqlite backends
- Use ssh as communication between frontend and hosts
- XML-RPC API
-
- Support for LVM contributed by CERN
- Enables *Hybrid clouds* (i.e instantiation on remote cloud providers)
- Implements subset of EC2 interface as well as upcoming OCCI interface for *Public cloud* interface.

OpenNebula.org

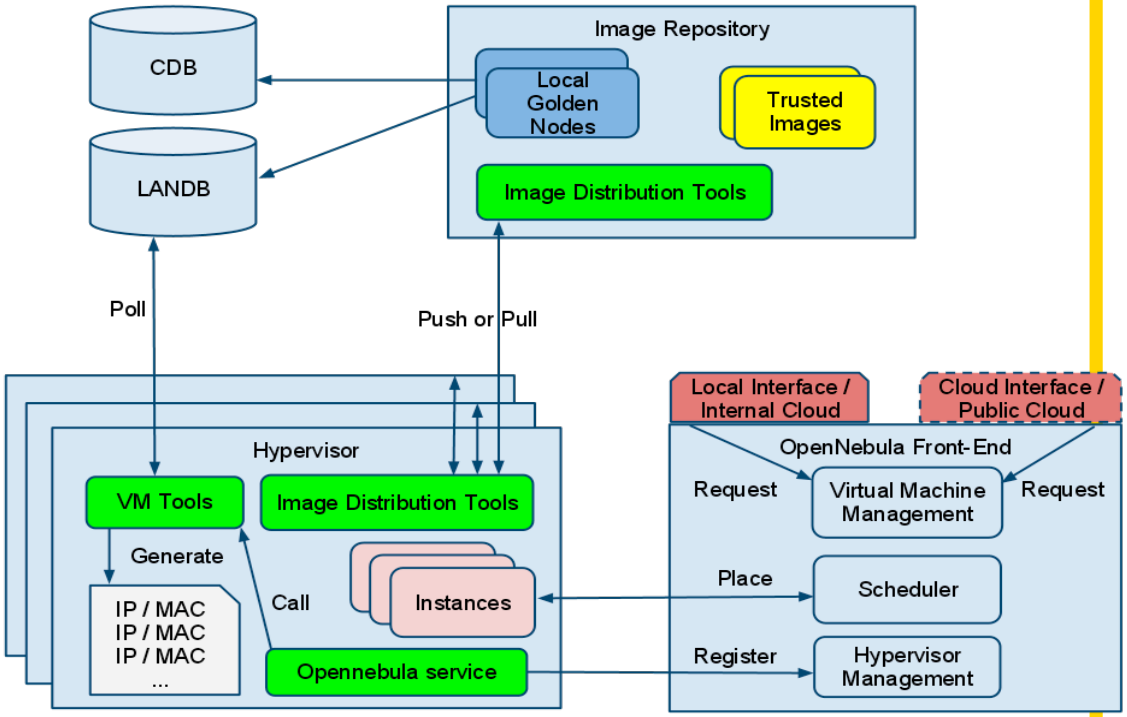
Comparison with Similar Technologies

OpenNebula - Architecture, Current Status & Roadmap

	Platform ISF	VMware Vsphere	Eucalyptus	Nimbus	OpenNebula
Virtualization Management	**VMware, Xen**	**VMware**	**Xen, KVM**	**Xen**	**Xen, KVM, VMware**
Virtual Network Management	**Yes**	**Yes**	No	**Yes**	**Yes**
Image Management	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Service Contextualization	No	No	No	**Yes**	**Yes**
Scheduling	**Yes**	**Yes**	No	No	**Yes**
Administration Interface	**Yes**	**Yes**	No	No	**Yes**
Hybrid Cloud Computing	No	No	No	No	**Yes**
Cloud Interfaces	No	**vCloud**	**EC2**	**WSRF, EC2**	**EC2 Query OGF OCCI vCloud**
Flexibility and Extensibility	**Yes**	No	**Yes**	**Yes**	**Yes**
Open Source	No	No	**GPL**	**Apache**	**Apache**

CERN's LXCLOUD details

- A log scp and bittorrent image distribution has been implemented
- Hypervisors run utilities to detect what VM they are allowed to run and which images they need to download
- OpenNebula triggers instantiation via ssh
- Instances based on LVM snapshots



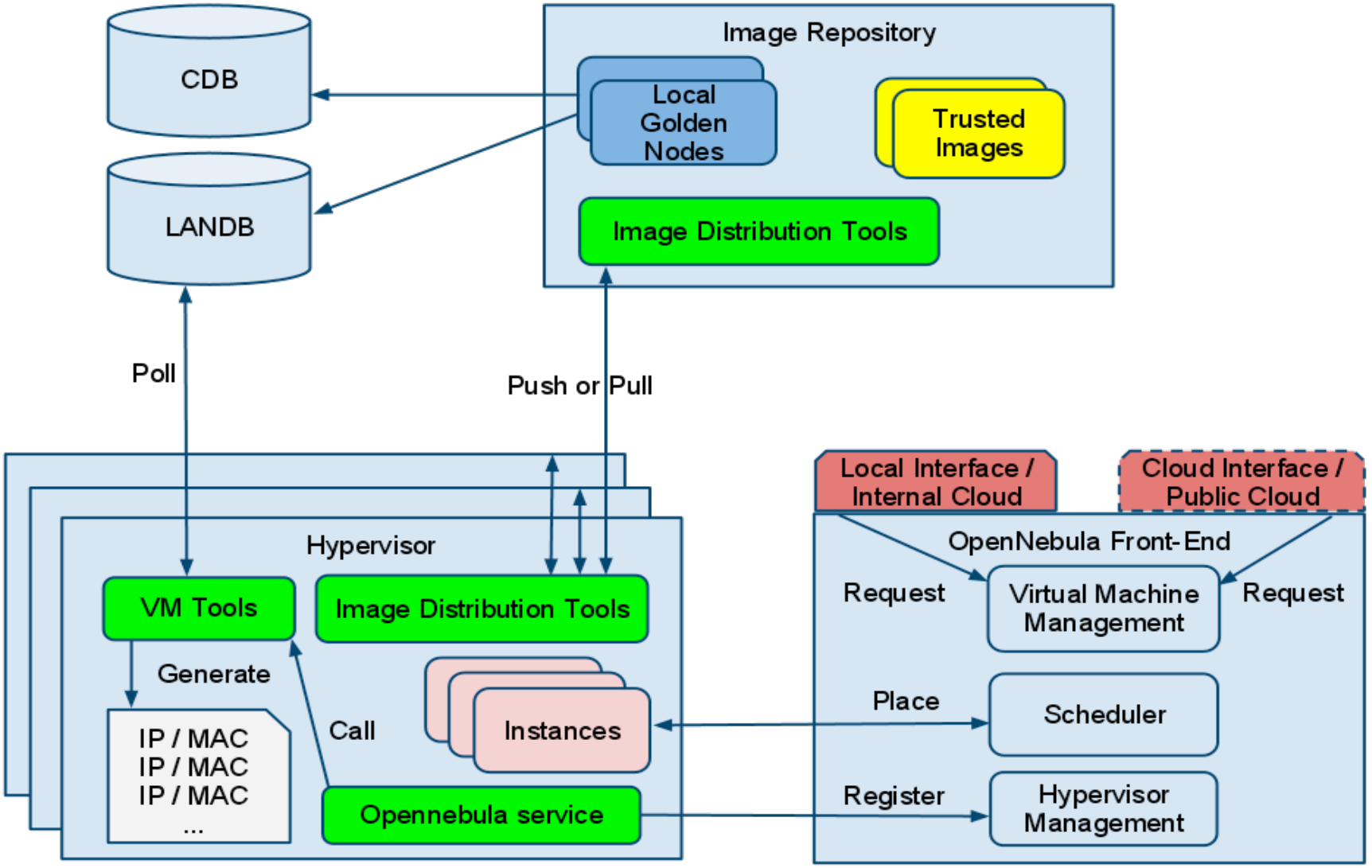

CDB
LANDB
Image Repository
Local Golden Nodes
Trusted Images
Image Distribution Tools
Poll
Push or Pull
Hypervisor
VM Tools
Image Distribution Tools
Generate
IP / MAC
IP / MAC
IP / MAC
...
Call
Instances
Opennebula service
Local Interface / Internal Cloud
Cloud Interface / Public Cloud
OpenNebula Front-End
Request
Virtual Machine Management
Request
Place
Scheduler
Register
Hypervisor Management

Image Distribution

Push:

- Sequential SCP
- logarithmic SCP (scp-wave)
- http://code.google.com/p/scp-wave/

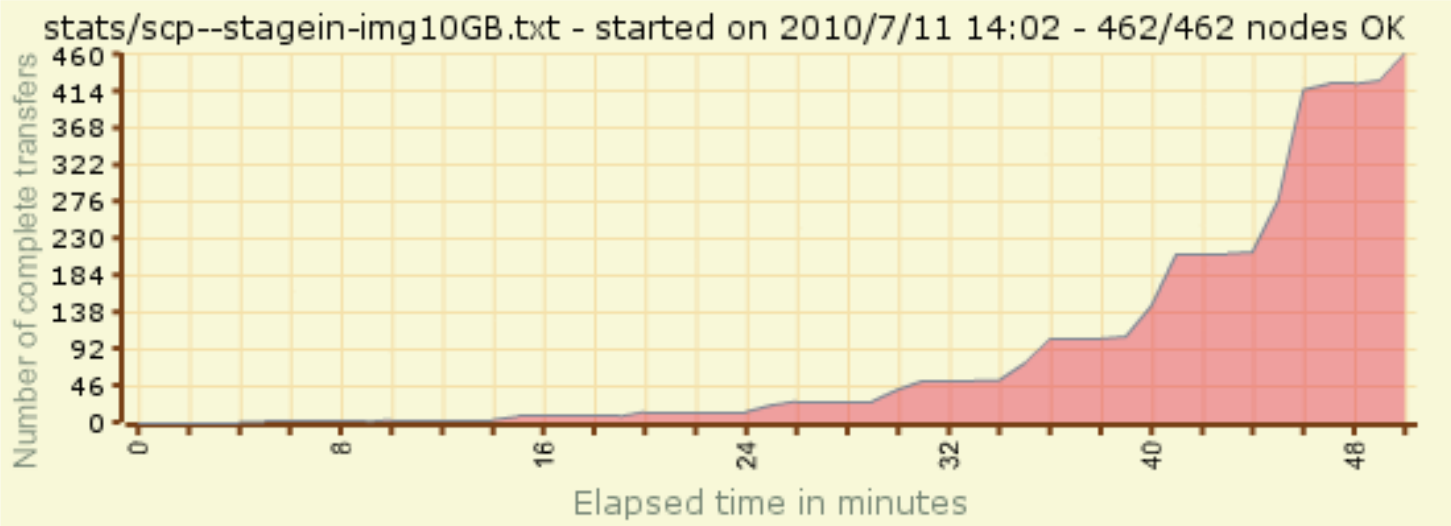
Pull:

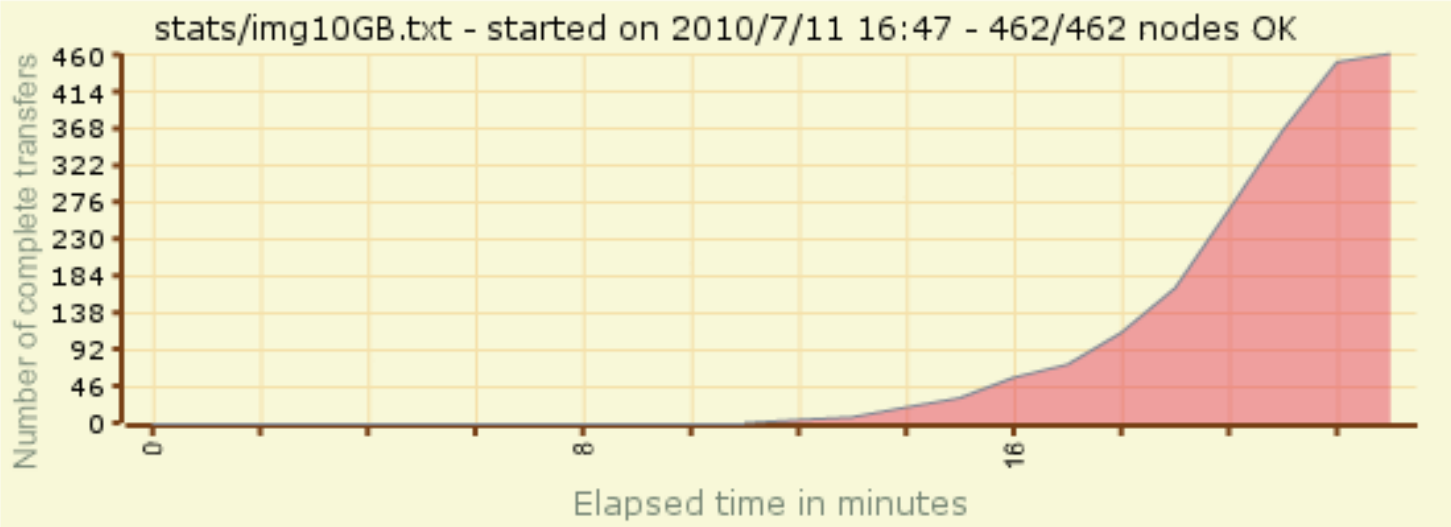
- wget via an http based repository (locally)
- Bittorrent (Romain Wartel, Belmiro Moreira @CERN)

Shared FS

- NFS
- PVFS, Lustre...

Image distribution results (thx to Belmiro)



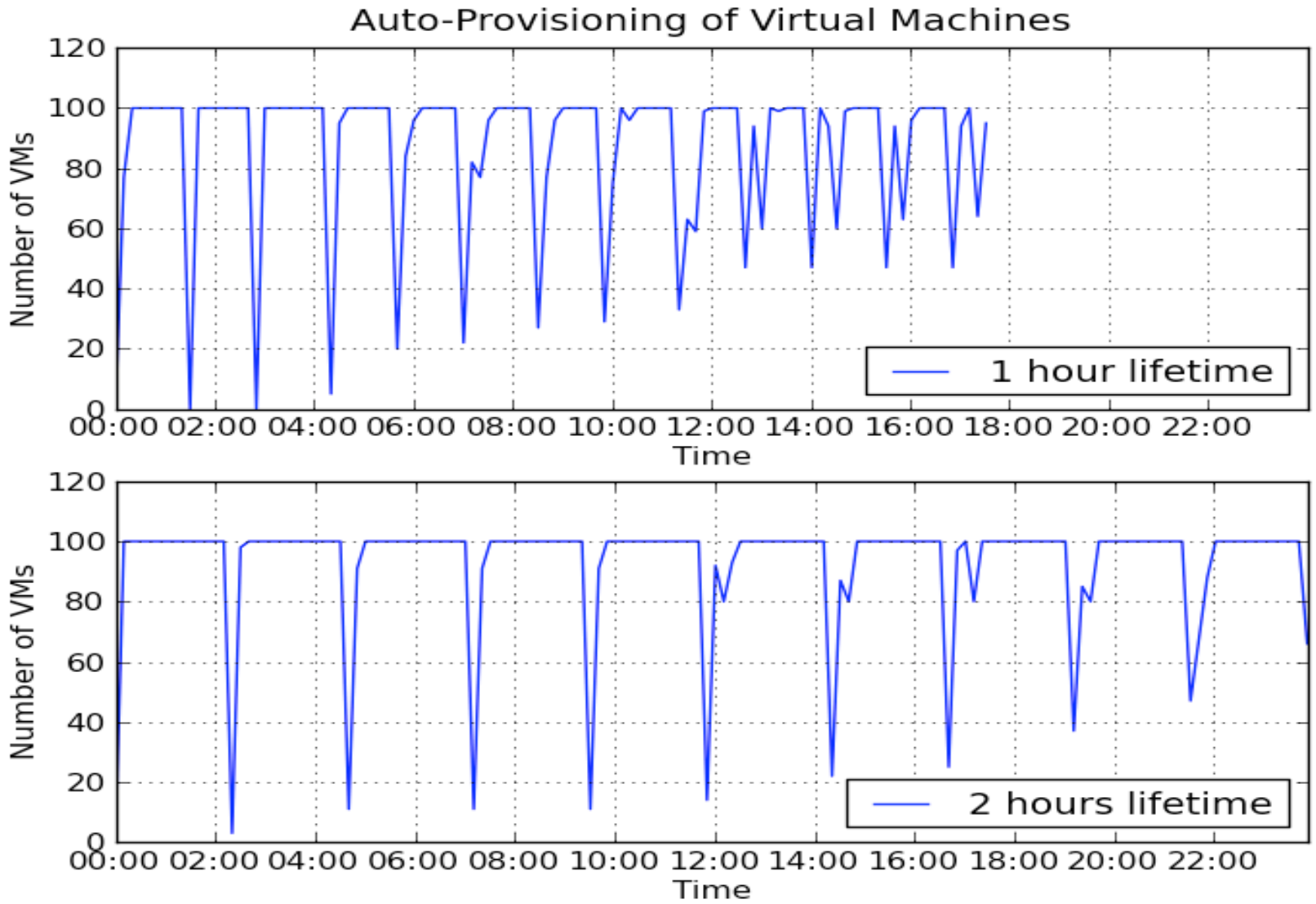


Guiding the provisioning

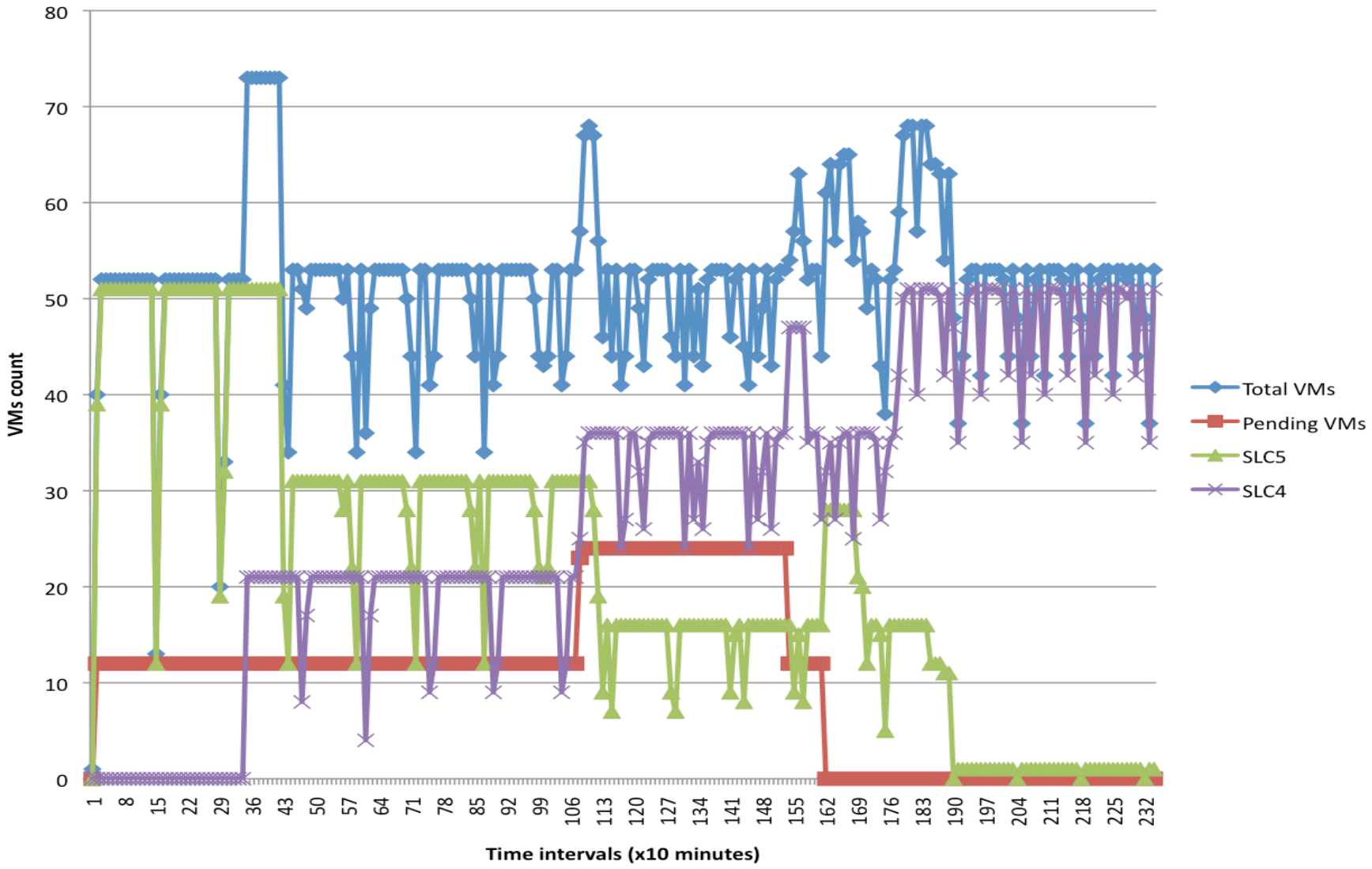
- Define policies to compose the batch farm
- Automate the provisioning of the virtual machines such that the policies are enforced.
- e.g Inspect the job queue and deduce the best composition of the batch farm. In terms of SMP VMs, OS...

- A sizer is used to monitor the pool of VM instances and evaluate the policies.
- Currently only one policy: "*Keep the pool full with the proper shares of VM types*"
- See ICAC2010 and CCGRID2009 papers

Autonomic Provisioning Results



Early Results of *sizer*



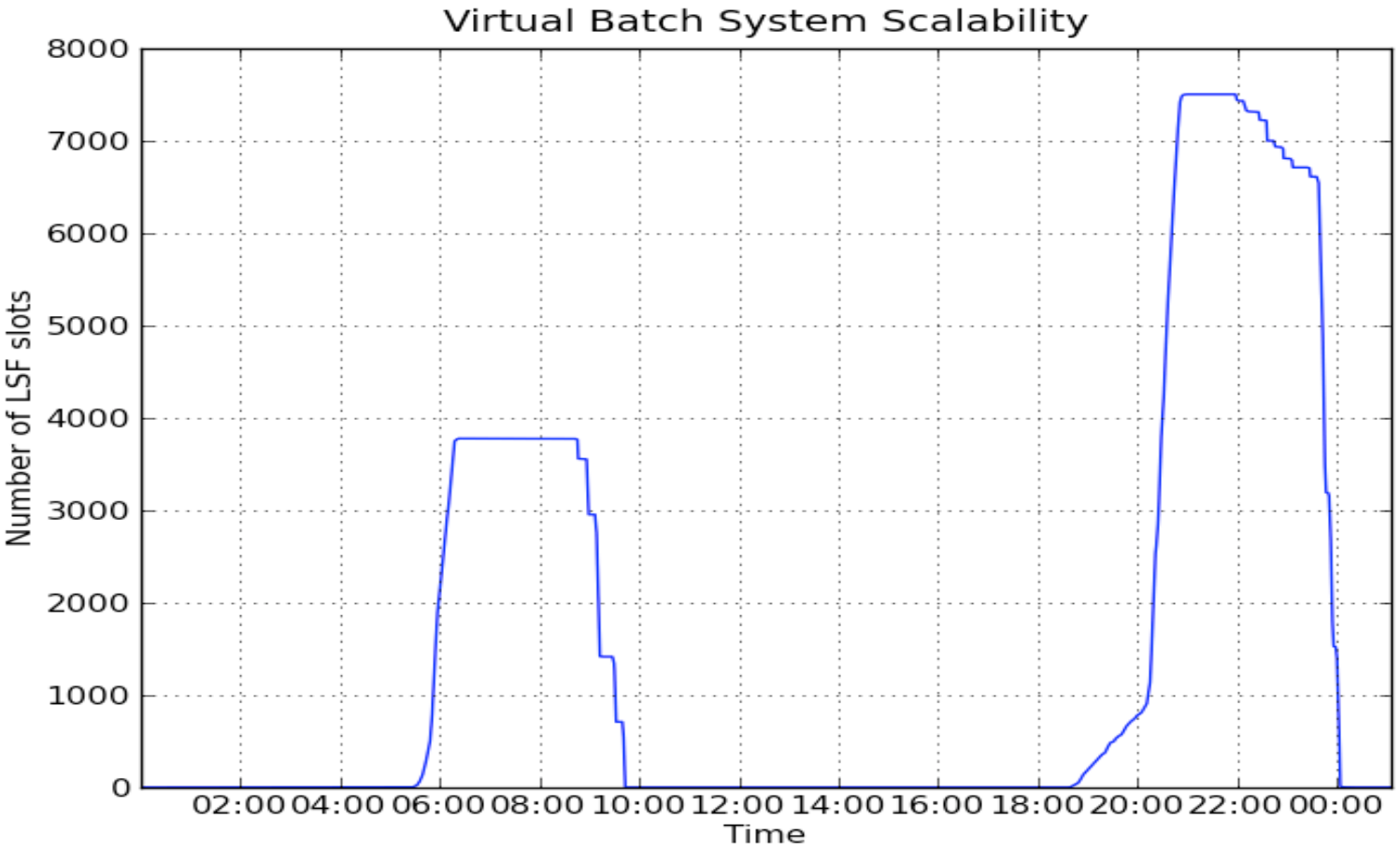
Joining the Batch system

...A contextualization problem...

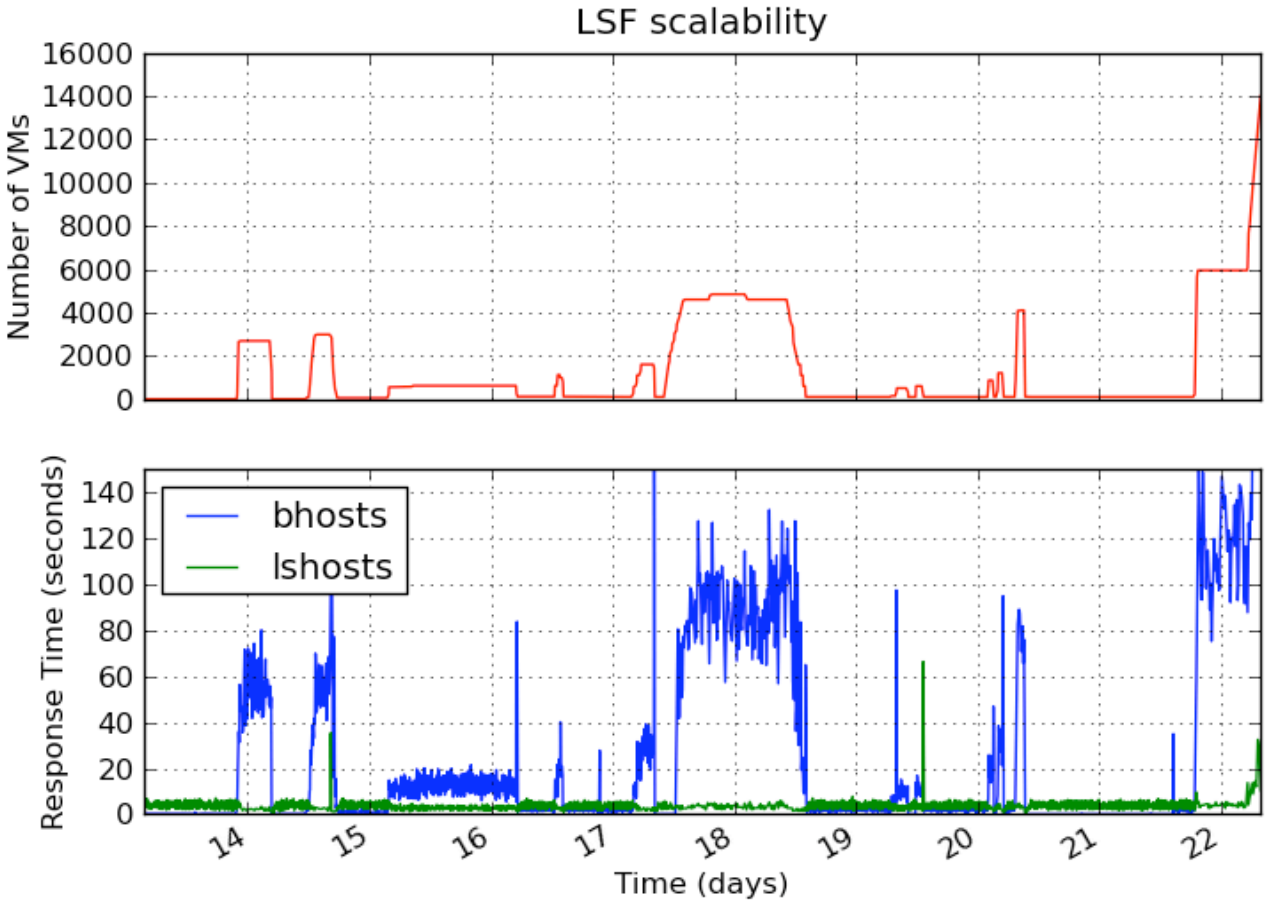
```
CONTEXT = [
vmid        = "$VMID",
TTL         = "3",
AFS         = "off",
files       = "/opt/vmimage/init.sh /opt/vmimage/etchosts /opt/vmimage/
etcsysconfigifcfg /opt/vmimage/id_rsa.pub /o
pt/vmimage/lsfcontext.conf /opt/vmimage/etcsysconfignetwork",
target      = "xvdb"
]
```

- Files and variables are stored in a ISO created on the fly.
- Startup script mounts this ISO and runs contextualization script.
- VMs are setup as dynamic hosts in the LSF pool.

Scalability Tests...7,500 slots in LSF via Opennebula



Testing LSF scalability



IaaS at Clemson

- The really easy way:
  - KVM on a regular HPC cluster
  - NAT networking (every VM gets its own NAT)
  - Base image on NFS server
  - KVM snapshot mode creates temporary disk in scratch, disk discarded once instance is shutdown
  - Submit VMs as PBS jobs

```
IMAGE=/home/sebgoa/kvm/star5.img
export TMPDIR=/local_scratch

kvm -hda $IMAGE -net nic,model=e1000 -net user -m 1280 -snapshot -nographic;
```

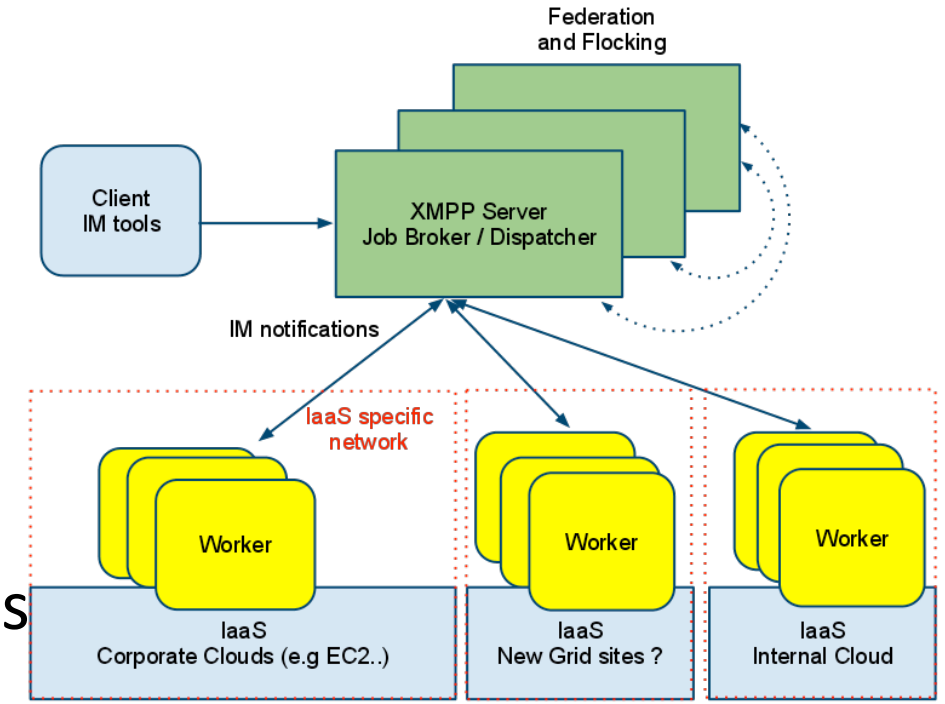
IaaS at Clemson

- But…
  - No shared FS between VMs
  - Looks like each VM has the same IP
  - Can't use regular job management systems to run jobs in those VMs (need glidein/proxy like solution)

- This setup has been one of the key drivers for our development of Kestrel: An XMPP based job management system

Kestrel

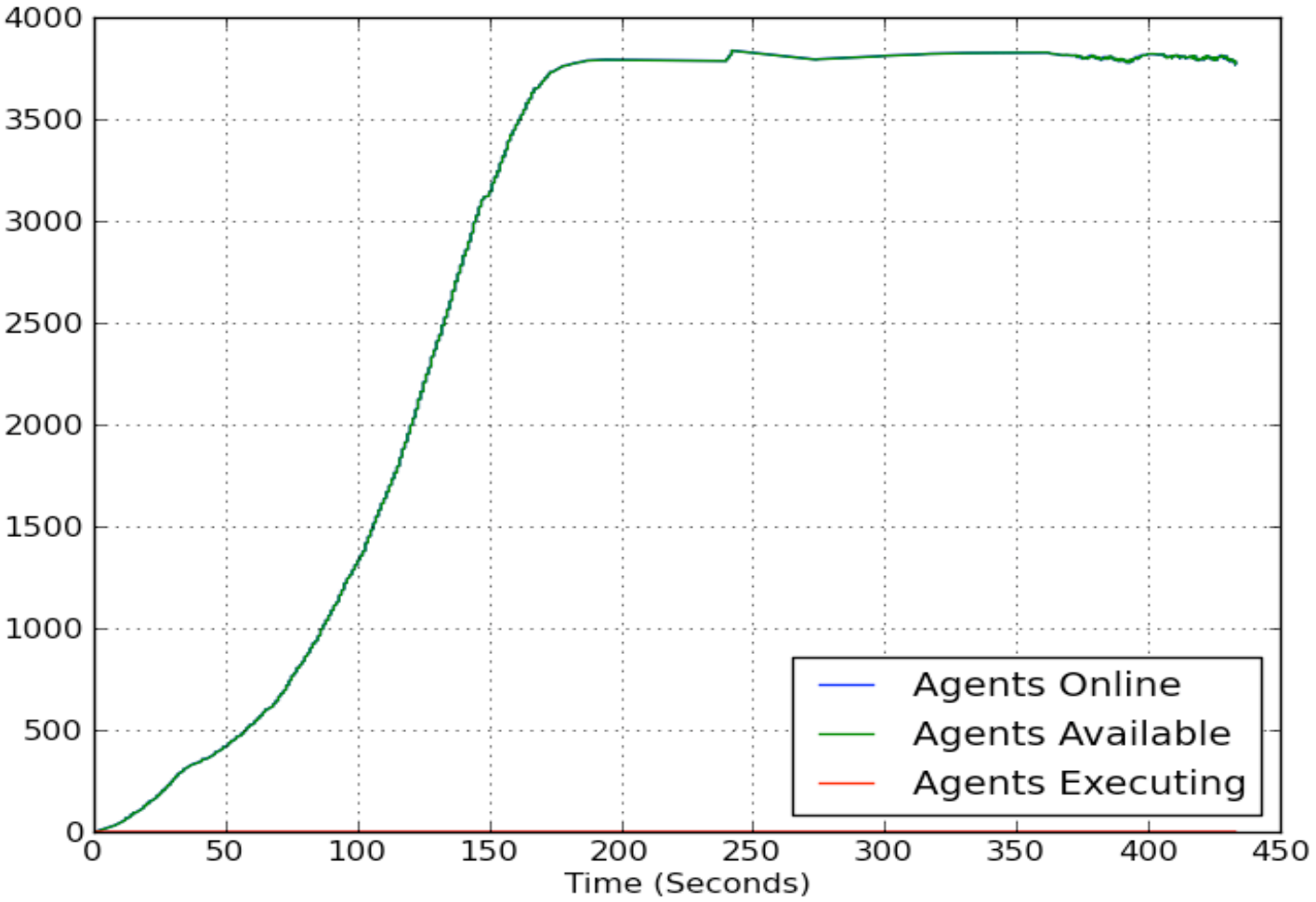
- A job management framework using the XMPP protocol
- Started as a student project
- Uses Instant Messaging concepts of notifications
- Practical in adverse network conditions



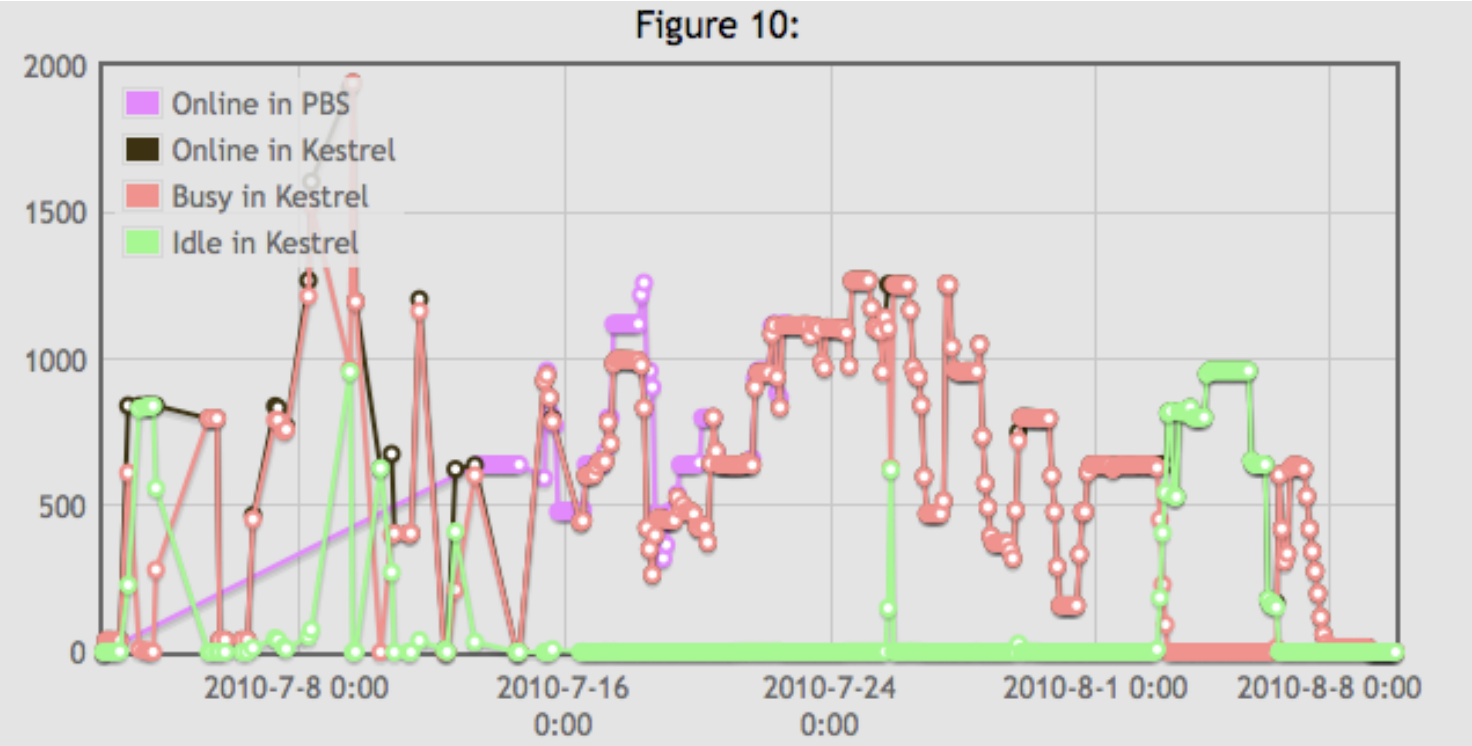
http://wiki.github.com/legastero/Kestrel/

https://twiki.grid.iu.edu/bin/view/CampusGrids/InstallingKestrel

Booting VMs is extremely fast (20 VMs/sec)



STAR Success with Clemson IaaS and Kestrel



- "But to simulate the equivalent sample of 12.2 Billion Monte-Carlo events with ~ 10 Million accepted by event triggering after full event reconstruction, we would have taken 3 years at BNL on 50 machines This Monte-Carlo event generation would essentially not have been done. With the resources from cloud, we took 3-4 weeks." –Jerome Lauret BNL.

Conclusions

- The Cloud is here let's hope it gets sunny
- API explosion opens up possibilities
- Focusing on IaaS layers, LXCLOUD and Clemson's clusters have been developed/enhanced to provision VMs.
- Great scalability with OpenNebula
- KVM shows great promise especially with the snapshot mode
- Performance will get even better
- May need specialized job management systems to make use of Clouds across multi-site

CLEMSON
SCHOOL OF COMPUTING

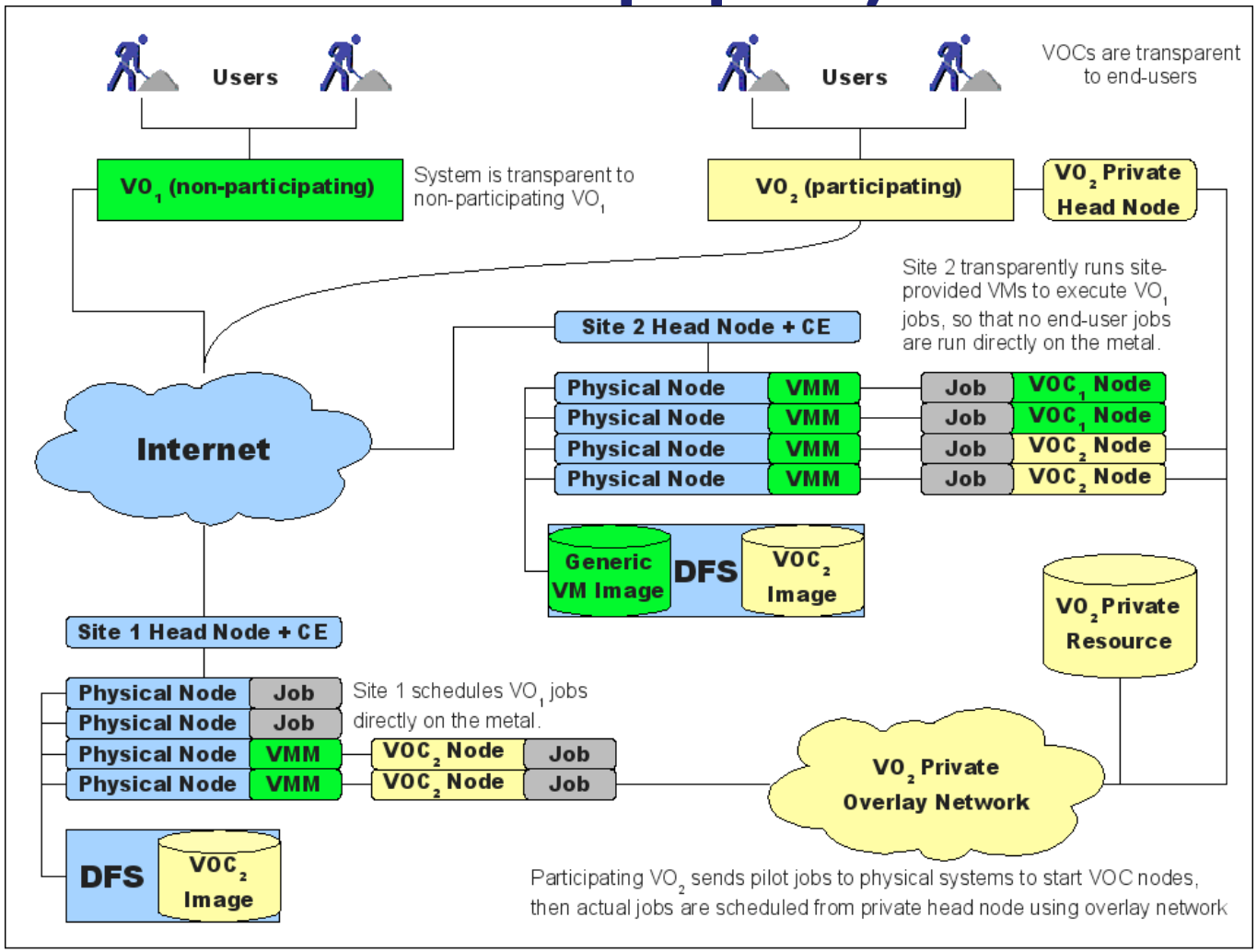
Thanks to NSF, DOE and OSG
Thanks to Lance Stout, Mike Murphy, Michael Fenn, Linton Abraham and all the other students…
Thanks to CERN and the IT/PES-PS group
Thanks to Jerome Lauret, Matthew Walker

Questions?: sebgoa@clemson.edu
http://cirg.cs.clemson.edu

Outline

- Cloud Basics
- Building a Cloud Provider
  - Lxcloud @ CERN (In collaboration with Ulrich Schwickerath, Ewan Roche, Belmiro Moreira and Romain Wartel)
- **VOCs and Clouds**
  - **Research done at Clemson**

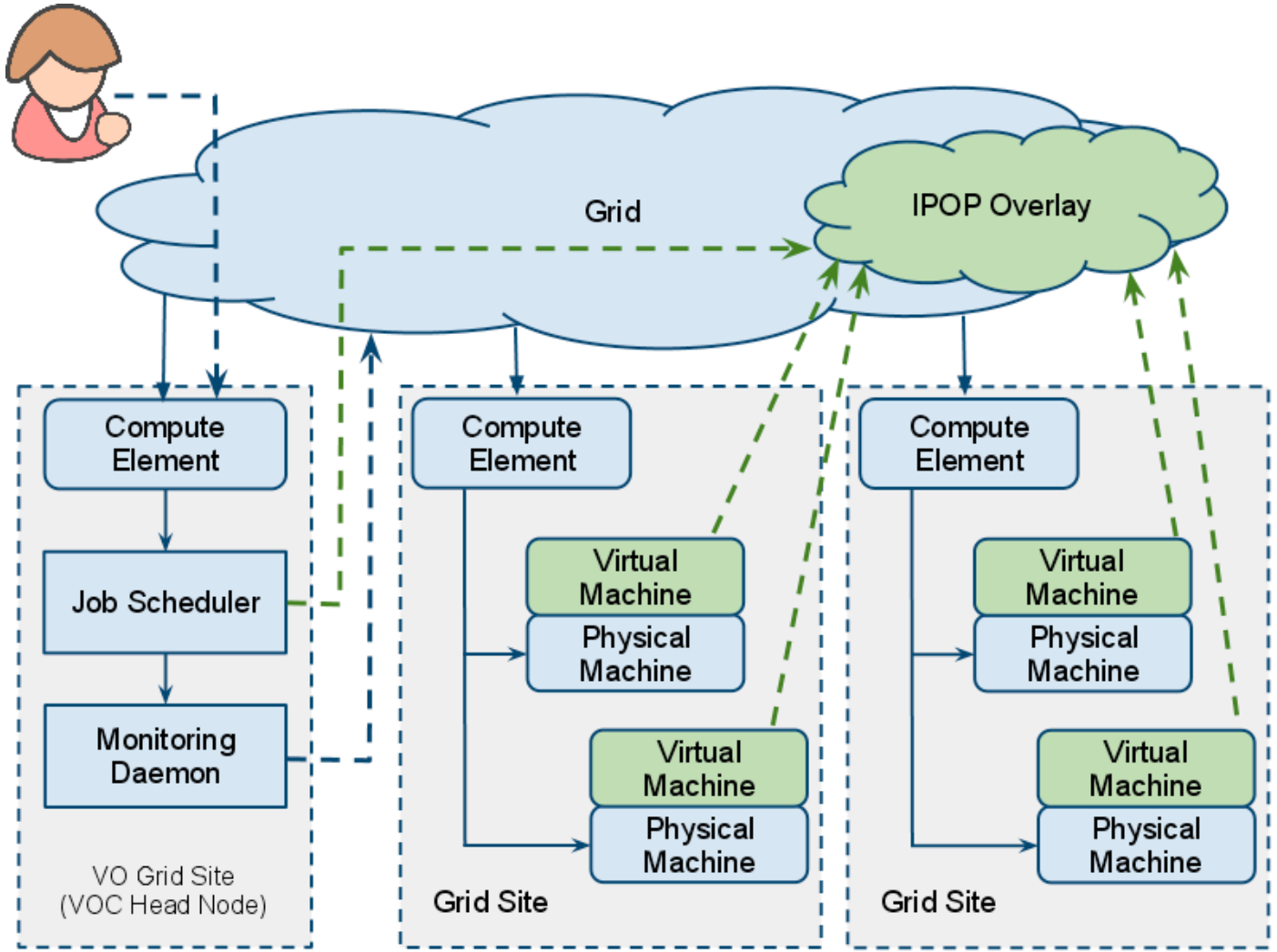
VOC: Virtual Organization Cluster (JGC +FGCS papers)



Why VOCs aka Clouds ?

- Observation that what people want is resources with their own OS/Apps and central scheduling: Pilot job frameworks.
- A cloud is a cluster over WAN
- Therefore there is a need for
  - A way to request/start the nodes
  - A way to create a virtual network
  - A way to run jobs in them
- Very similar to glideinWMS but the pilots ask to start VMs

Multi-Site Overlay (ICAC 2010)



VOC Implementation

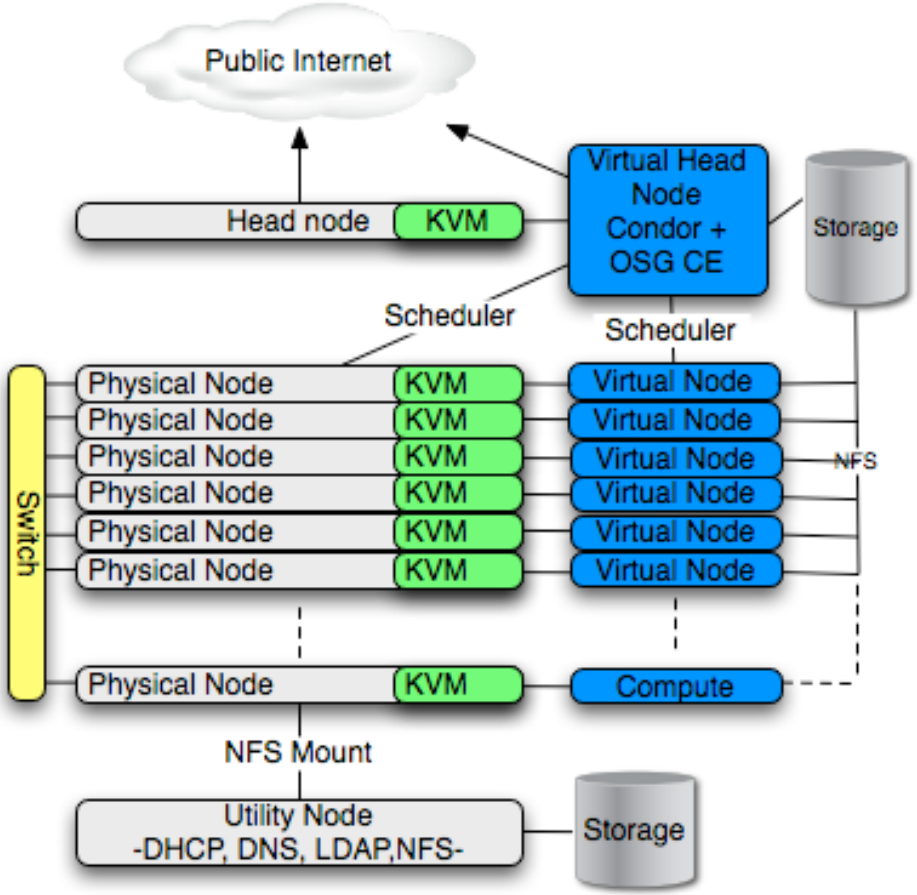
- Multiple configurations:
  - **Type 1:** Shared head node on Physical cluster, VO is unaware of VOC (e.g LXCLOUD)
  - **Type 2:** VO provides virtual head nodes on multiple grid sites.
  - **Type 3:** VO uses an overlay network with a single head node (e.g STAR).

Type 1: Implementation

- KVM vs. Xen for ease of use
- Normal Cluster utilities/ techniques

- NFS share
- And PVFS setup

- KVM offers a snapshot mode that gives us ability to use a single image file. Writes are temporary

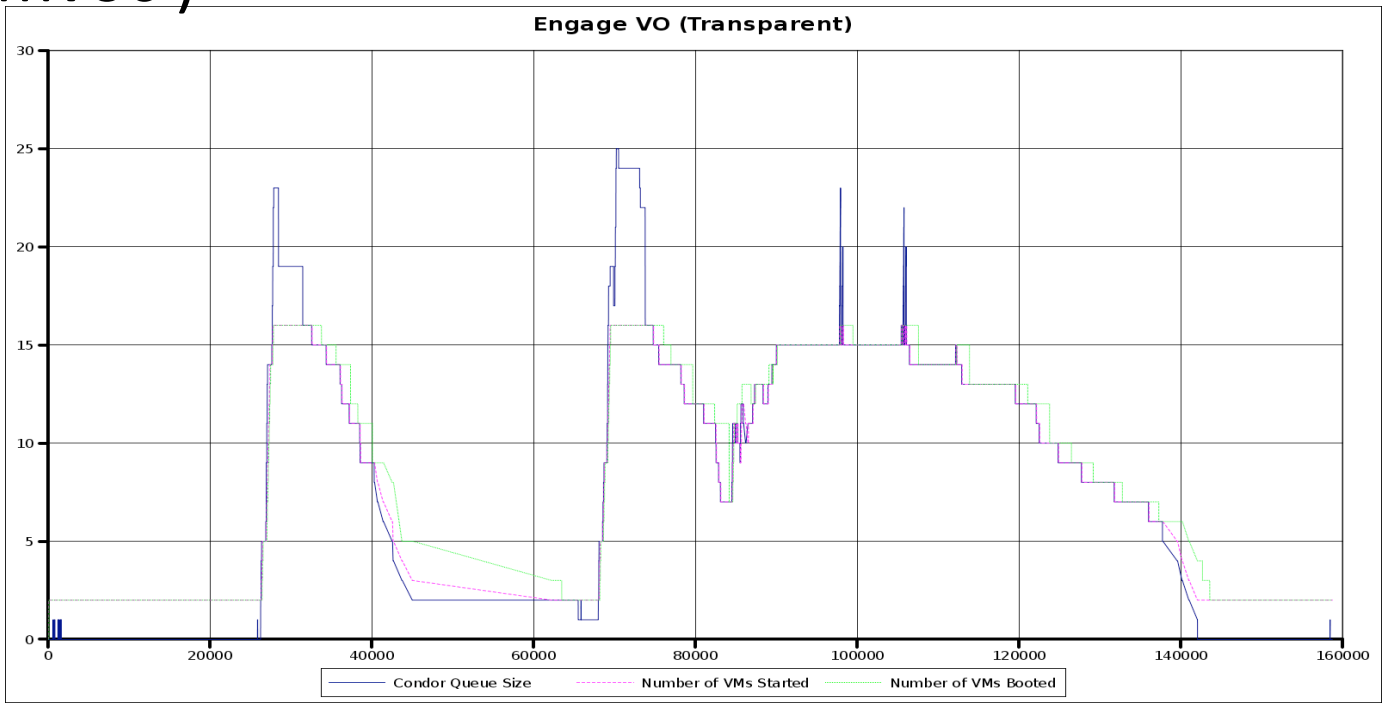


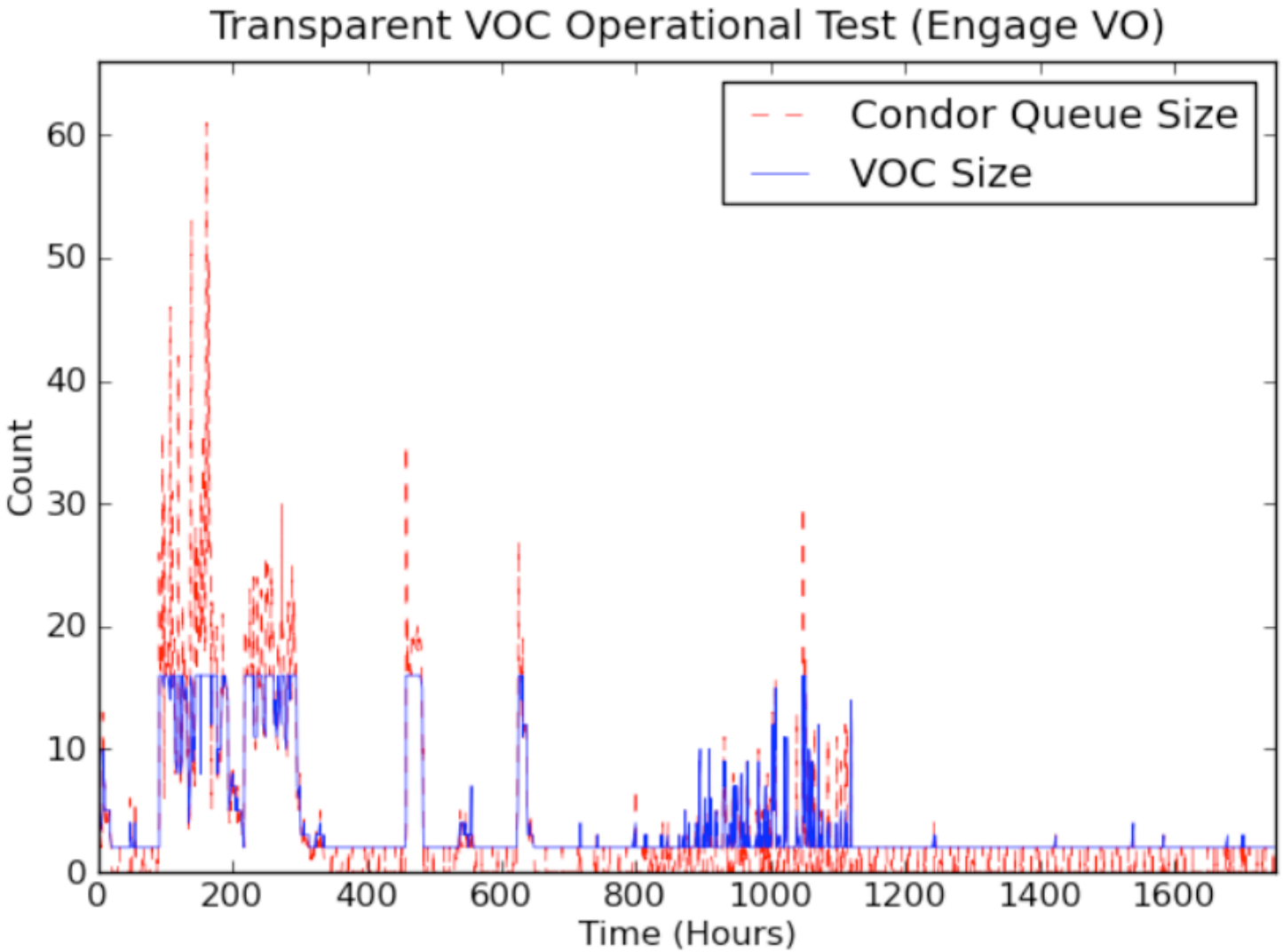
Load-Driven Provisioning (CCGRID09)

- **Dynamic Provisioning is done via the use of a Watchdog on the VOC head node**
- Watchdog monitors incoming jobs on the OSG gatekeeper (Condor job manager is used)
- When jobs are in the local scheduler queue, the watchdog starts a VM on a physical host (static mapping between host and guest currently). XML-RPC system
- When VM starts, Condor inside the VM starts and advertizes its presence to the central manager -> Jobs run.

Experimental Results

- Engage VO on OSG
- Site Clemson-Birdnest on OSG Production
- Cluster size responds to load, Simulation Results confirm (Pending IPDPS paper, simulator: simVOC available at http://cirg.cs.clemson.edu/software/simvoc)





Engage VO on OSG

From: ACAT 2010, February 22-27th Jaipur/ India